STRUCTURAL ANALYSIS OF DNA REPLICATION ACROSS UNSTABLE REPETITIVE SEQUENCES

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Cindy Follonier

von

Vernamiège VS

Promotionskomitee

Prof. Massimo Lopes (Vorsitz und Leitung der Dissertation)

Prof. Joachim Lingner

Prof. Toshio Mori

Prof. Peter Beard

Prof. Urs Greber

Zürich, 2012

Table of content

Zusammenfassung

Die Verkürzung oder Expansion von sich wiederholenden Trinukleotidesequenzen, sogenannten „Trinukleotid-repeats“ (TNR), ist die Ursache für neurodegenerative Krankheiten wie Friedreichs Ataxie (GAA), die Huntington-Krankheit (CAG) oder das Fragile-X-Syndrom (CGG). Lange TNR Sequenzen können alternative DNS-Sekundärstrukturen *in vitro* bilden und hemmen das Fortschreiten von DNS Replikationsgabeln in Hefezellen und Bakterien. In menschlichen Zellen sind die molekularen Mechanismen, die die DNS Replikation beeinträchtigen und zur Expansion der TNR führen, allerdings weitgehend unbekannt. Wir haben ein experimentelles System etabliert, um die *in vivo* Replikationsstrukturen („replication intermediates“, RI) zu analysieren, die bei der DNS Replikation von GAA-Trinukleotidsequenzen entstehen. Dabei transfizieren wir humane Zellen mit Plasmiden, die GAA-Sequenzen in unterschiedlichen Längen und Orientierungen enthalten. Nach Replikation dieser Plasmide in den transfizierten Zellen isolieren wir die RI und analysieren sie mittels bidimensionalen (2D) Agarosegeln und dem Elektronenmikroskop (EM).

Unsere 2D-Gel-Analysen von RI aus humanen 293T und U2OS Zellen zeigt, dass Replikationsgabeln durch GAA-Sequenzen nur transient angehalten werden, und dass dieser Effekt von der Länge und Orientierung der TNR abhängt. Zu unserer Überraschung haben wir ausserdem noch weitere Signale in unseren 2D-Gelen erhalten, deren Auftreten mit der Länge von TNR, bei der Symptome von Friedreichs Ataxie (FRDA) auftreten, korreliert. Mit Hilfe des EM haben wir sowohl die gesamte RI Population, als auch die Moleküle, die wir durch Elution der genannten Signale aus unseren 2D-Gelen isoliert haben, umfassend analysiert. Dabei haben wir erstmals hoch aufgelöste Bilder der Strukturen gewonnen, mit denen Schwesterchromatiden unmittelbar hinter der Replikationsgabel miteinander verbunden sind. Bei ungestörter Replikation sind diese Verbindungen willkürlich über die gesamte Länge der replizierten Moleküle verteilt. Im Gegensatz dazu führen expandierte GAA-Sequenzen zu einer Stabilisierung dieser Verbindungen in der repetitiven Sequenz. Darüber hinaus führen GAA-Sequenzen zur Reversion der Replikationsgabel *in vivo* und

beeinflussen gleichzeitig die Stabilität der zweiten Replikationsgabel des Replikons. Die Ergebnisse unsere Experimente legen nahe, dass postreplikative Strukturen für die GAA-Triplettexpansion und damit für das Auftreten von Friedreichs Ataxie verantwortlich sind. Ähnliche Vorgänge könnten ursächlich für die Expansion andere TNR-Sequenzen sein, die mit einer wachsenden Zahl neurodegenerativer Erkrankungen in Verbindung gebracht werden.

Die experimentelle Identifikation an der Expansion von GAA-Sequenzen beteiligter zellulärer Faktoren und die Entwicklung effektiver Diagnosetechniken sind bisher durch methodische Schwierigkeiten bei der Detektion expandierter TNR eingeschränkt. Für eine effektive Diagnose und ein tieferes Verständnis der molekularen Grundlagen der FRDA sind die schnelle und zuverlässige Detektion expandierter TNR aber Voraussetzung. Ausgehend von isolierter DNS mit GAA-Sequenzen und den damit verbundenen alternativen Strukturen haben wir in Zusammenarbeit mit der Gruppe von Dr. Toshio Mori einen Antikörper etabliert, der spezifisch DNS Epitope in expandierten GAA-Sequenzen erkennt. Unsere *in vitro* Experimente haben die Spezifität dieses Antikörpers bestätigt, aber auch gezeigt, dass eine Detektion von GAA-assoziierten Strukturen *in vivo* aufgrund des hohen Überschusses normaler DNS mit diesem Antikörper nicht möglich ist. Daher haben wir uns auf die Verfeinerung unserer *in vitro* Techniken konzentriert, um das analytische Potential dieses Antikörpers optimal auszunutzen und zusätzliche Informationen über den Einfluss von GAA-Sequenzen sowohl auf die DNS Replikation als auch auf die Transkription zu gewinnen.

Summary

Trinucleotide repeats (TNR) can undergo large deletions or expansions, leading to neurodegenerative diseases like Friedreich's ataxia (GAA), Huntington disease (CAG) or Fragile X (CGG). Expanded TNR have been long suggested to form *in vitro* non-B secondary structures and were shown to stall DNA replication forks in yeast and bacteria. However, the molecular mechanisms leading *in vivo* to DNA replication interference and repeat expansion in human cells have remained elusive. We established an experimental system to analyse *in vivo* DNA replication intermediates (RI) across GAA repeats in human cells. SV40-derived plasmids containing GAA repeats in different number and orientation are transfected into human cells and allowed to replicate. Plasmid RI are recovered and analyzed by a combination of bidimensional (2D) agarose gels and electron microscopy (EM).

Our 2D-gel analysis in 293T and U2OS cells reveals that replication forks are only transiently paused by GAA repeats in a length- and orientation-dependent manner. Surprisingly, besides fork pausing, additional unexpected signals are observed, which correlate with the minimal number of GAA repeats known to cause symptoms in Friedreich's ataxia (FRDA) patients. Extensive EM analysis of the total RI population, as well as the molecules eluted from defined position of the 2D gels, allowed us for the first time to visualize at high resolution how sister chromatids are connected shortly behind a moving fork during unperturbed replication. Importantly, while in control conditions these junctions happen randomly along the replicated duplexes and do not involve homologous sequences, expanded GAA repeats stabilize sister chromatid pairing directly at the TNR. We have also found that GAA repeats induce fork reversal *in vivo* and affect the integrity of the sister fork in the replicon. Altogether, these results suggest that postreplicative structures could be a molecular determinant of GAA expansion, leading to Friedreich's Ataxia. Similar mechanisms could be involved in the expansion of other TNR, linked to a growing number of human neurodegenerative syndromes.

Some limitations of the analytical methods currently employed to monitor GAA expansion have hampered the development of effective diagnostic tools and the identification of novel cellular factors involved in TNR expansion. Both effective diagnosis and molecular understanding of FRDA onset await the development of a fast and functional read-out for the consequences of GAA expansion. Using isolated GAA-associated non-B DNA structures as an antigen and taking advantage of an intercontinental collaboration with the highly specialized group of Prof. Toshio Mori, we established an antibody able to specifically recognize DNA epitopes associated with expanded GAA repeats. Further *in vitro* characterisation of this antibody confirmed its specificity, but also revealed its inability to detect GAA secondary structure when embedded in an excess of competing DNA, impairing possible *in vivo* application. We have started to explore more sophisticated *in vitro* experiments, where this new molecular tool could provide additional insight on the interference of GAA repeats with replication and transcription.

1. Introduction

All the genetic information essential for life is encoded by the DNA. In order to maintain this code from generation to generation the DNA has to be copied by a mechanism called replication. The proteins involved have to face different problems or damages caused by external factors, like UV irradiation, or by internal impediments, like repetitive sequences or secondary structures. To maintain the accuracy of replication cells developed different strategies, like polymerase proofreading activity and post-replication repair mechanisms. The combination of replication and repair prevents too frequent modifications of the genome, which would lead to disease.

1.1. Cell cycle

The division of a mother into two identical daughter cells follows a highly controlled cycle, which is named the cell cycle. It is composed of 4 different stages with specific functions. The first one is the G1 phase. It is the phase during which the cell takes the time to grow after cell division and decides or not to enter a new cycle. If the environment is not good or if its genome is damaged, the cell can exit the cell cycle until the conditions get better.

Once the cell passes this checkpoint control, it enters in S-phase, where the DNA will be replicated. After duplication, the integrity of the genome is assessed by a second checkpoint control during G2/M phase. Different repair mechanisms collaborate to solve damages and check if the DNA is fully replicated. The accuracy of S and G2 phases is very important to prevent the inheritance of mutations in the new generation. When all the checkpoint conditions are fulfilled the cell can finally divide, by a mechanism called mitosis, giving rise to two identical daughter cells (M-phase).

1.2. Eukaryotic DNA replication

DNA replication is a complex mechanism essential for the survival of cells and organisms. Any mistake during this process can lead to severe consequences. In order to prevent mutations, genome duplication is tightly regulated. Replication takes place in 3 steps: initiation, elongation and termination.

The initiation step starts in G1 with the binding of the origin recognition complex ORC1-6, Cdc6, Cdt1 and Mcm2-7 helicases to the origin of replication forming the pre-replication complex (Fig1a) (Blow and Hodgson, 2002). At the end of G1 and during S phase the assembly of pre-replication complexes is inhibited, in order to prevent re-licensing, which would lead to re-replication. Once the cells enter S-phase Mcm2-7 loaded origins are activated and recruit the proteins forming the replication fork (RF) complex (Fig1b).

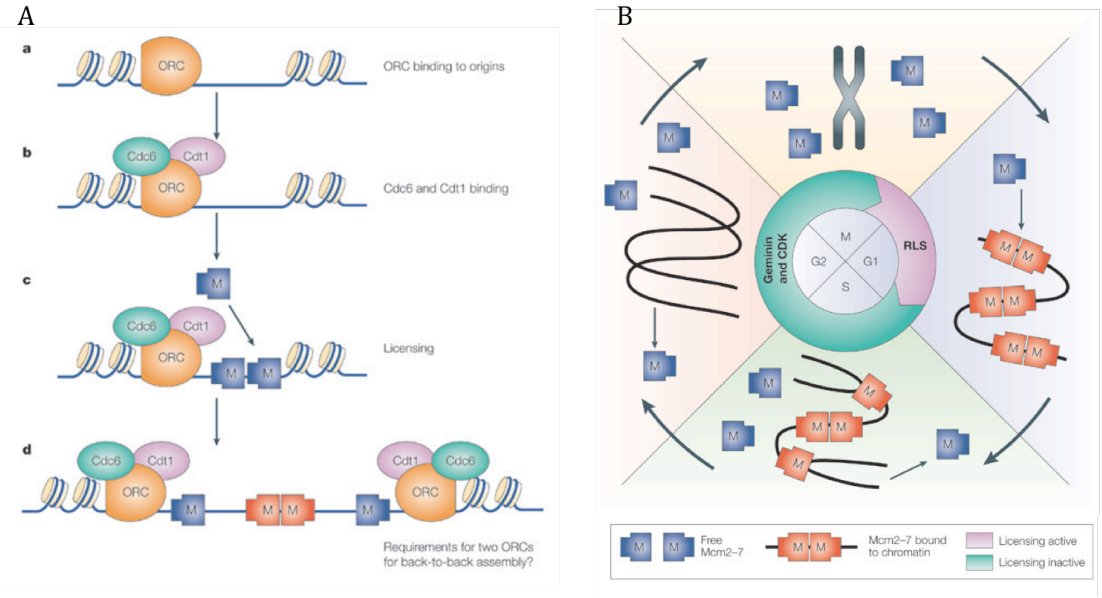


Figure 1. A. Origin licensing. a) The origin recognition complex (ORC) is first recruited to the replication origin; b) ORC recruits Cdc6 and Cdt1; c) ORC, Cdc6 and Cdt1 act together to load Mcm2–7 protein hexamers onto the origin, which licenses the DNA for replication; d) Initiation-competent complexes. B. Regulated loading and unloading of Mcm2–7 during the cell cycle. (Modified from (Blow and Dutta, 2005)

Each origin generates two sister RFs, which move into opposite directions (Fig2). In order to progress, they first need to unwind the parental DNA using the Mcm2-7 helicase. Once the DNA is unwound the synthesis of the daughter strands can take place. The proteins in charge of this step are the DNA polymerases (pol). They can add deoxyribonucleotides (dNTP) to a DNA strand, but only in a 5' to 3' orientation. This directionality specificity implies the distinction between the leading strand, where the DNA is synthesized in a continuous manner, and the lagging strand, where the synthesis is discontinuous. Moreover, the DNA polymerases cannot polymerise *de novo*; they require the presence of a pre-existing 3' end. This is the role of polα primase, a polymerase migrating with the helicase, which is able to synthesize short RNA sequences, - called primers -, on the DNA, which serve as starting points for DNA polymerisation.

In the case of the leading strand, the polα primase synthesizes one RNA primer at the time of origin firing from which polε can elongate. As the polymerization takes place in the same direction as the helicase, polε can copy the DNA in one single stretch (Pursell et al., 2007). Besides its 5'-3' polymerase activity, polε has also a proofreading 3'-5' exonuclease activity, which reduces the risk of fixing mutations. The lagging strand has a more complicated replication process. Due to the 5'-3' activity of the polymerase, this strand is elongated in the direction opposite to the helicase, requiring a stepwise polymerization mechanism. First, the helicase unwinds the parental double helix, releasing free single strand DNA (ssDNA), which is directly covered by a single strand binding protein called RPA. RPA protects the ssDNA from re-annealing or breakage until it is replicated. The primase synthesises an RNA primer from which DNA polδ can elongate. Once the polymerization has started the RNA primer can be removed by RNaseH (Balakrishnan et al., 2010). Finally, when polδ reaches the previous Okasaki fragment, the two newly synthesized DNA fragments are sealed by ligaseI forming one continuous DNA strand.

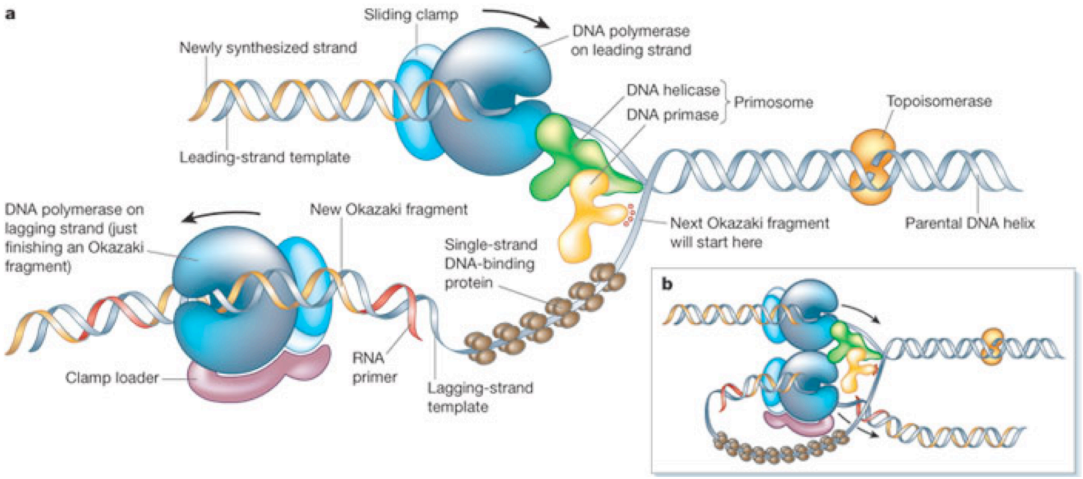


Figure 2. A. Replication fork proteins. B. 3D representation of the replication fork. (Alberts, 2003)

When replication is completed within a certain genomic area, RFs meet each other and fuse. This step is called termination. The exact mechanism leading to fork fusion and complete duplication of the genome is still unclear and object of intense study.

1.3. Replication fork stabilization at a pausing site

1.3.1. Replication pausing complex and post-replicative repair

During replication, the forks have to face different kind of obstacles. The first ones are endogenous, like long repetitive sequences, secondary structures or interference with concomitant DNA metabolism mechanisms, such as transcription. Additional challenges come from external factors that cause DNA modifications, such as irradiation (e.g. UV light), methylating agents (e.g. methylmethasulfonate, MMS) or nucleotide depletion (hydroxyurea, HU). Under these specific circumstances it is particularly important to maintain fork integrity and to effectively restart the fork after repair.

To prevent RF to dissociate while meeting impediments, different proteins not directly involved in DNA synthesis are required. Among them are TIM, TIPIN and CLASPIN, which form the so-called "replication pausing complex" (Errico and Costanzo, 2010)(Fig3). This complex travels with the RF in presence or absence of DNA damage and interacts with the helicase (Chou and Elledge, 2006) and polε (Sercin and Kemp, 2011). In case of damage, CLASPIN activates the ATR-Chk1 cascade (Tanaka, 2010), which helps maintaining fork stability and assisting the repair mechanisms.

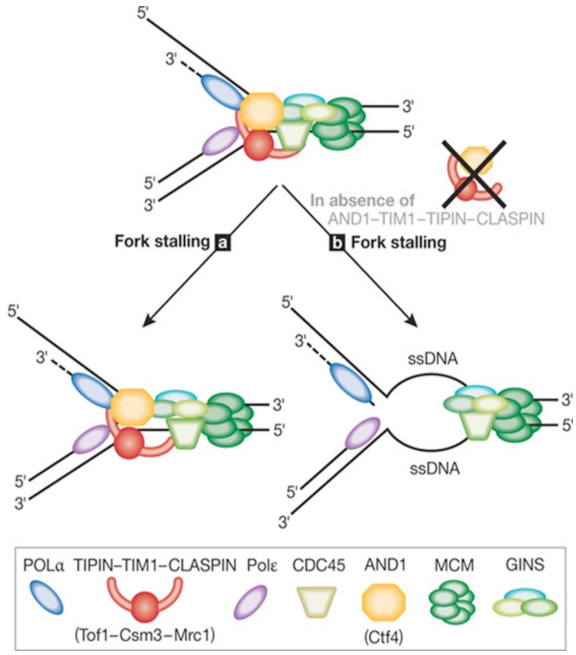


Figure 3. A. When replication is halted, TIPIN–TIM1 and CLASPIN physically link the polymerase and helicase activities, preventing fork collapse. B. The absence of these components could lead to excessive unwinding of DNA, thus destabilizing the replisome. (Errico and Costanzo, 2010)

Lesions can also occur behind the replication fork and block the polymerases. To solve such damages a specific repair mechanism is activated, the post-replicative repair (PRR). PRR can be divided in two different pathways, the translesion synthesis and the template switch. Translesion synthesis recruits specific polymerases to bypass the lesion, whereas the template switch is based on recombination-like mechanisms directly at the fork (fork reversal) or behind the fork.

1.3.2. Fork reversal

In the 1970s Higgins et al. (Higgins et al., 1976) identified by EM a new replication intermediate. It consists of a 4-way junction, which forms upon DNA damage. BrdU incorporation and CsCl gradient data suggested that the 4th arm arises from the annealing of the two newly synthesized strands at the replication fork. Based on these results and some former publications about strand displacement (Masamune and Richardson, 1971) and branch migration (Broker and Lehman, 1971), a new model for DNA damage tolerance during replication was proposed (Fig4), nowadays known as “fork reversal”. In presence of damaging agents, DNA synthesis proceeds until one of the polymerases is blocked by a lesion, while the second can continue past the damage (Fig4A). This step can be followed by the reassociation of the parental strands by branch migration inducing the regression of the two nascent strands, which anneal with each other (Fig4B). The blocked polymerase can then use the newly synthesized strand as a template, bypassing the damage present on the parental strand (Fig4C). The reversed replication fork can then be remodeled to a normal fork and directly restart without removing the initial lesion (Fig4D).

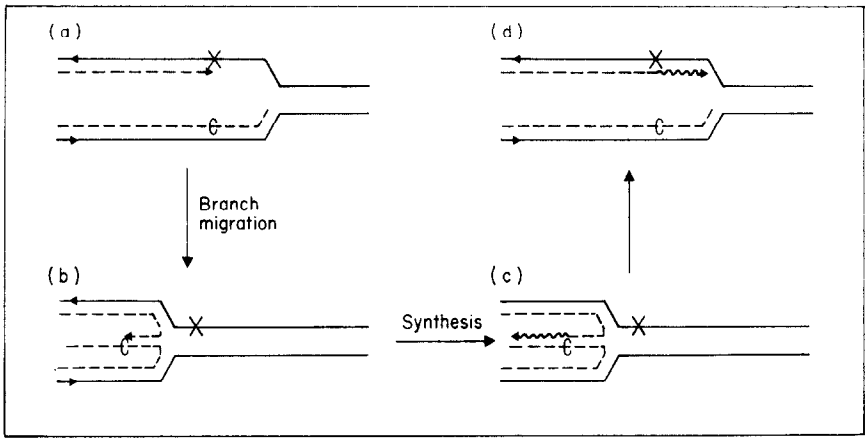


Figure 4. Model of replication repair. Strand displacement and branch migration create an alternate template allowing replication to bypass a lesion (X). (Higgins et al., 1976)

This model was largely ignored for almost 20 years, until new evidence supported its biological relevance. Data in *E.coli* (Flores et al., 2001) showed that blockage of the lagging strand synthesis induces the formation of a Holliday-like junction at the fork, which can be processed by the RuvABC helicase/endonuclease, in the absence of the recombination factor RecF. To explain these results the author referred to the fork reversal model. In yeast, Holliday-like structures were also observed by EM at replication forks, when checkpoint deficient cells were challenged by nucleotide depletion (HU; Sogo et al., 2002). As similar structures were not detected in wild type cells, fork reversal was described mainly as a pathological phenomenon.

Since then, growing evidence supported the occurrence of fork reversal, but the factors inducing it and its biological relevance have remained elusive. Specialized helicases seem to be required for fork regression. *In vitro* assays show that the higher eukaryotic helicases of the RecQ family (BLM and WRN), as well as bacterial RecG, have the potential to catalyse the regression of model fork structures (reviewed in Atkinson and McGlynn, 2009). Despite the absence of helicase activity, Rad5, a yeast protein genetically linked to the template switch pathway, has the potential to induce regression and branch migration of the fork *in vitro* (reviewed in (Atkinson and McGlynn, 2009).

Bacterial data also suggest a role of DNA topological stress: increased supercoiling is sufficient to induce fork regression (Postow et al., 2001). Recent

unpublished data (Ray Chaudhuri and Lopes, manuscript submitted) demonstrate that fork reversal is a frequent event in yeast, Xenopus and mammals upon camptothecin (CPT) treatment. CPT is an inhibitor of the topoisomerase I, which prevents the release of DNA supercoiling in front of the fork, potentially leading to fork stalling (Koster et al., 2007). Overall, experiments ongoing in the lab are accumulating evidence of fork reversal in response to several different challenges to the replication process - even in presence of a functional DNA damage checkpoint - (unpublished data from Arnab Ray Chaudhuri and Kai Neelsen) suggesting fork regression as a general replication fork stabilization mechanism, rather than as a pathological structure.

1.4. Post-replicative junctions

2D-gel analysis of replication intermediates in Xenopus (Lucas and Hyrien, 2000), Physarum (Benard et al., 2001) and yeasts (Segurado et al., 2002); (Lopes et al., 2003) showed during DNA replication the formation of post-replicative junctions, which migrate as a 2n-spike. Characterization of these intermediates indicates that they are X-shaped molecules formed of two fully replicated sister chromatids. Xenopus (Lucas and Hyrien, 2000) and yeast (Lopes et al., 2003) studies revealed that the junction of the Xs is sequence independent and can branch migrate irrespective of Mg^{+2} consistent with the presence of an hemicatenane structure. However, a direct demonstration of the molecular architecture of this junction was missing and represents one of the major achievements of this PhD work.

Postreplicative junctions are also detected upon replication of a damaged template. Although the exact mechanism by which these post-replicative junctions are formed still needs to be clarified, they have been genetically linked to the template switch pathway and they have been considered closely related to the damage-independent hemicatenanes (Branzei et al., 2008). Under normal conditions these junctions are resolved by RecQ helicases together with the topoisomerase III (Liberi et al., 2005). Based on these results a model combining template switch pathway and hemicatenane formation was proposed (Liberi et al., 2005) (Fig5): when the replication polymerase stalls at the damage, the

hemicatenane branch migrates behind the fork and unwinds the two newly synthesized strands from their template inducing template switch (Fig5B). The final intermediate will look like a double Holliday junction (Fig5C), which can be resolved by RecQ helicase and topIII (Fig5D). In absence of hemicatenane (formerly resolved by a nick, for example) the template switch might be induced by a recombination-dependent mechanism, which is supposed to act in parallel or as a back up pathway (Branzei; Liberi et al., 2005).

The biological function of these post-replicative junctions behind the fork could have different roles, like maintaining the two sister chromatids together to favour the deposition of cohesin rings, or keeping a broken chromatid next to its sister strand to assist repair, or supporting the detection of sequence homology to promote faster recombination. However, none of these hypotheses have been confirmed yet, due to the absence of direct evidence proving the existence of hemicatenanes.

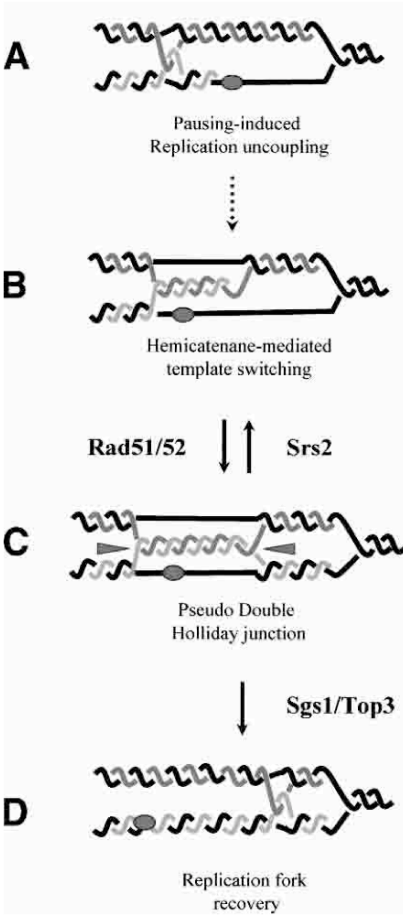


Figure 5. A model for DNA damage replication bypass. (Liberi et al., 2005)

1.5. DNA repeats in regulation and disease

1.5.1. Repeats forming secondary structures involved in regulation mechanisms

Repetitive sequences are frequent in the genome and they often have the capacity to form non-B DNA structures (Fig6). Firstly considered only as pathologic, some of these repeats were involved in different regulatory mechanisms. Thus RFs consistently have to deal with secondary structures. If they are not resolved before replication, they can interfere with the replication process and cause genomic instability. For this reason the cell evolved different mechanisms to prevent such problems. An example of repetitive sequences forming secondary structures and involved in DNA metabolism is described below.

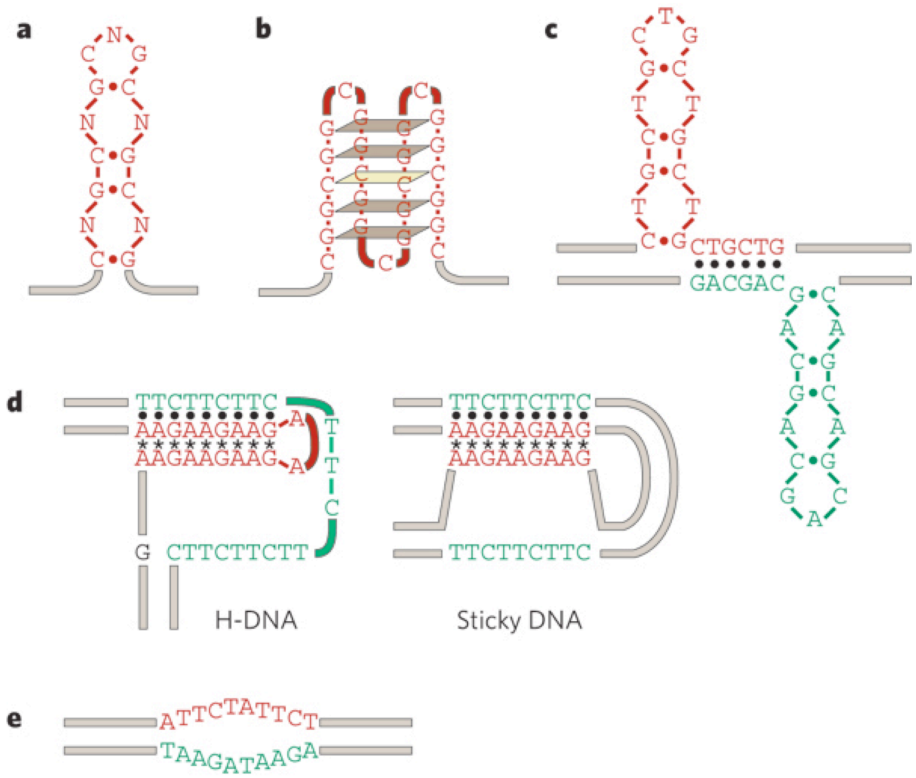


Figure 6. Repetitive DNA can form several unusual structures, examples of which are shown. The structure-prone strand of the repetitive run is shown in red, its complementary strand in green, and flanking DNA in beige. a, An imperfect hairpin formed by $(CNG)_n$ repeats. b, A quadruplex-like structure formed by the $(CGG)_n$ repeat. c, A slipped stranded structure formed by the $(CTG)_n$•$(CAG)_n$ repeat. d, H-DNA and sticky DNA formed by the $(GAA)_n$•$(TTC)_n$ repeat. Only one possible isoform, in which the homopurine strand is donated to the triplex, is shown for both structures. Reverse Hoogsteen pairing is indicated by asterisks. e, A DNA-unwinding element formed by the $(ATTCT)_n$•$(AGAAT)_n$ repeat. (Mirkin and Mirkin, 2007)

The most studied type of repeats are the telomeres. Located at the extremity of the chromosomes, they consist of mutliple repetitions of TTAGGG sequence and can measure between 3-15kb. Due to the high content of guanines, telomeres have the ability to form G-quadruplexes (G4) (Sampathi and Chai, 2011) (Fig6b). These are stable inter- or intra-molecular structures stabilized by Hoogsteen base-pairing.

Genetics studies showed that different helicases (WRN, BLM) have the potential to solve G4 structure, improving replication fork progression at telomeres (Chavez et al., 2009). In their absence the fork stalls and a fast and lethal shortening of the telomeric sequences is observed (Chavez et al., 2009). These data clearly show that DNA replication and secondary structure formation have to be tightly regulated in order to prevent damages.

Guanine rich sequences with the potential to form G-quartet are not specific to telomeres. They have also been identified in other regions of the genome indicating that replication indeed has to frequently face this kind of impediment. Computational analyses show that >40% of the promoters contain 1 or more quadruplex motif (Huppert and Balasubramanian, 2007) suggesting its implication in gene regulation. An increased amount and stability of such structure was also observed in the 3-UTR regions, implying a role in transcriptional termination (Huppert et al., 2008).

1.5.2. Repeats forming pathological secondary structures: TNRs

The reason why repetitive sequences were firstly considered as pathologic comes from the discovery that a large number (>30) of neurodegenerative diseases are due to microsatellite expansion and that an important increase of mono- and dinucleotide repeats was detected in certain human cancers (Mirkin and Mirkin, 2007).

One subgroup of the microsatellite family is specifically associated with rare neurological diseases: the trinucleotide repeats (TNR). Each pathology is related to one specific triplet, like CAG for the Huntington chorea and Spinocerebellar ataxia, CGG for the fragile X, GAA for Friedreich ataxia or CTG for the myotonic dystrophy (Table1)(McMurray, 2010). The symptoms appear when the triplets expand over a certain threshold, which is different for each pathology. However,

before getting sick the patients pass through a pre-mutation phase, during which the repeats start to expand, but are still too short to cause symptoms. The phenotypes observed for each disease are not only due to the sequence of the repeats, but also to their ability to form secondary structures and to their location. To illustrate these points, examples are described in the following paragraphs.

Table 1. Features of TNR repeat expansion in humans. (McMurray, 2010)

Disease	Sequence	Location	Parent of origin of expansion	Repeat number (normal)	Repeat number (pre-mutation)	Repeat number (disease)	Somatic instability
Diseases with coding TNRs							
DRPLA	CAG	*ATN1* (exon 5)	P	6–35	35–48	49–88	Yes
HD	CAG	*HTT* (exon 1)	P	6–29	29–37	38–180	Yes
OPMD	GCN	*PABPN1* (exon 1)	P and M	10	12–17	>11	None found in tissue tested (hypothalamus)
SCA1	CAG	*ATXN1* (exon 8)	P	6–39	40	41–83	Yes
SCA2	CAG	*ATXN2* (exon 1)	P	<31	31–32	32–200	Unknown
SCA3 (Machado–Joseph disease)	CAG	*ATXN3* (exon 8)	P	12–40	41–85	52–86	Unknown
SCA6	CAG	*CACNA1A* (exon 47)	P	<18	19	20–33	None found
SCA7	CAG	*ATXN7* (exon 3)	P	4–17	28–33	>36 to >460	Yes
SCA17	CAG	*TBP* (exon 3)	P > M	25–42	43–48	45–66	Yes
SMBA	CAG	*AR* (exon 1)	P	13–31	32–39	40	None found
Diseases with non-coding TNRs							
DM1	CTG	*DMPK* (3′ UTR)	M	5–37	37–50	<50	Yes
DM2	CCTG	*CNBP* (intron 1)	Uncertain	<30	31–74	75–11,000	Yes
FRAX-E	GCC	*AFF2* (5′ UTR)	M	4–39	40–200	>200	Unknown
FRDA	GAA	*FXN* (intron 1)	Recessive	5–30	31–100	70–1,000	Yes
FXS	CGG	*FMR1* (5′ UTR)	M	6–50	55–200	200–4,000	Yes
HDL2	CTG	*JPH3* (exon 2A)	M	6–27	29–35	36–57	Unknown
SCA8	CTG	*ATXN8OS* (3′ UTR)	M	15–34	34–89	89–250	Unknown
SCA10	ATTCT	*ATXN10* (intron 9)	M and P (smaller changes with M)	10–29	29–400	400–4,500	Yes
SCA12	CAG	*PPP2R2B* (5′ UTR)	M and P (more unstable with P)	7–28	28–66	66–78	None found

Huntington's disease is caused by the expansion of CAG repeats. Triplets start to be unstable between 29-37 repetitions and over this number most of the patients develop symptoms (McMurray, 2010). The repeats are located in the coding sequence of the Huntingtin protein and close to the end of the gene. A large expansion of the repeats increases the number of glutamines at the N-terminus of the protein leading to its aggregation and malfunction (Zuccato et al., 2010). A correlation was found between the ability of CAG to form slipped-strand DNA (Fig6c) *in vitro* after denaturing and renaturing conditions and the development of the chorea (Pearson et al., 1998). These data suggest a role of secondary structure formation in the sudden and fast expansion of the triplet repeats.

Patients suffering of the Fragile X disease have an expansion of the CGG repeats in the promoter of the fragile X mental retardation 1 gene (FMR1). The increase of guanines and cytosines causes a hypermethylation of the promoter, leading to reduced expression of the FMR1 protein. A repeat number between 55-200 corresponds to the pre-mutation phase and the full mutation status is reached when the number of triplets expands over 200 (McLennan et al., 2011). Long CGG have also the potential to form hairpins (Fig6a) (Chen et al., 1995) or G-quadruplexes (Fig6b) (Fry and Loeb, 1994) *in vitro* and can stall the replication fork *in vivo* (Voineagu et al., 2009).

GAA repeats are related to a disease called Friedreich's ataxia. Contrary to the triplets described above, GAA tracts are located in the intron of the frataxin gene (Mirkin and Mirkin, 2007) and their expansion impairs the splicing of the RNA, reducing the production of the protein (Ohshima et al., 1998; Shishkin et al., 2009). Data have shown that long GAA repeats (>50) can form a secondary structure in bacteria, which has the characteristics of a triplex (Fig6d)(Sakamoto et al., 1999), and can stall the replication fork (Krasilnikova and Mirkin, 2004) (Shishkin et al., 2009). The detection of the secondary structure correlates with the appearance of the symptoms (Sakamoto et al., 1999).

These examples highlight the similarities between TNRs, but also their differences. Triplets, over a certain threshold, have the potential to expand, form secondary structures and stall the replication fork. However, the location, the threshold and the secondary structures are specific to each specific trinucleotide. The difficulty to detect repeat expansion and secondary structures *in vivo* limits the understanding of TNR instability and a few major questions still remain in the field. How do they expand? How do they interfere with replication? Are the secondary structures also formed *in vivo* in mammalian cells?
Many studies have already tried to address these questions, but it would be too long to study in detail each specific triplet. That is why the rest of this thesis will be focused only on GAA repeats.

1.6. GAA repeats

1.6.1. Friedreich's ataxia: the disease

Friedreich's ataxia is the most frequent hereditary Caucasian autosomal recessive disorder, 1:50,000 people are affected in that region. The symptoms (ataxia, sensory loss, muscle weaknesses…) usually appear at the puberty and lead to the need of a wheelchair and daily help after 10 to 15 years (Pandolfo, 2009). No efficient treatment is available at moment to cure this severe pathology. In Switzerland, about 200 persons are supposed to be affected by Friedreich's ataxia, meaning 1 out of 35,000 (no official statistics available, numbers coming from the Swiss Association of Friedreich Ataxia "aCHaf").

This disease is specific to Caucasians because it originates from a unique founder mutation, the expansion of GAA repeats, located on the chromosome IX in the frataxin (FXN) gene. This gene gives rise to a 1.3kb transcript containing 5 introns, which encodes for a 210 amino acid protein, the frataxin. FXN is directed to the mitochondria, where it is supposed to take part in the iron metabolism (Santos et al., 2010). Its precise cellular role has still not been determined, but patient cells show problems in respiration, iron–sulfur cluster assembly, iron homeostasis, and maintenance of the redox status (Santos et al., 2010).

1.6.2. Frataxin downregulation due to GAA repeats

The reduction of frataxin expression in patient cells is correlated to the length of the GAA repeats located in the first intron of the gene. Several studies have tried to elucidate how the expanded repeats affect the transcription of FXN gene.

In vitro transcription assays demonstrated that GAA repeats impair transcription elongation by T7 RNA polymerase in a length dependent manner (Grabczyk and Usdin, 2000b). The stalling of the polymerase depends on the formation of d(TTC).r(GAA) hybrid (Fig7), which anneals behind the RNA polymerase (Fig7) (Grabczyk and Usdin, 2000b). *In vivo* studies confirm the effect of GAA secondary structure on transcription. *E.coli* data show a correlation between the formation of DNA.RNA hybrids at the repeats and RNA polymerase stalling (Grabczyk et al., 2007). Publications in mammalian (Ohshima et al., 1998) or patient cells (Punga and Buhler, 2010) also prove that long GAA tracts reduce the transcription efficiency by impairing RNA polymerase progression.

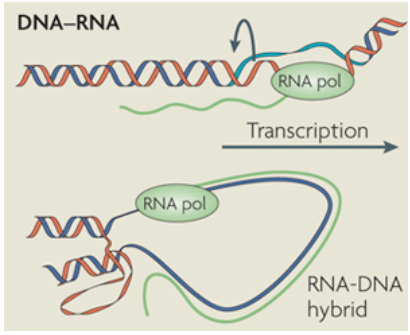


Figure 7. RNA.DNA hybrid formation during replication. Instead of being free the newly synthesized RNA reanneals behind the RNA polymerase. (Modified from Lopez Castel et al., 2010)

The reduction of frataxin protein can also be due to problems in the maturation of the transcript. Experiments done in yeast (Shishkin et al., 2009) and in COS1 cells (Baralle et al., 2008) prove that a long GAA tract in the intron of a gene can prevent the turn over of a pre-mRNA into a mature mRNA or cause the formation of aberrant splicing products, resulting in the down-expression of the protein.

1.6.3. Destabilisation of triplex formation can restore FXN transcription

The amelioration of frataxin transcription is one of the key targets to cure Friedreich's ataxia patients. A publication from 2000 (Grabczyk and Usdin, 2000a) showed *in vitro* that deoxyribonucleotides designed to prevent the formation or stabilization of triplexes can restore the transcription of a GAA-containing gene. Unfortunately such oligonucleotides cannot be used as therapeutics as they can not be delivered to the cells.

However, these results primed new attempts to identify a ligand that could reduce triplex folding and access the DNA in the nucleus. In one of these studies 126 compounds were tested on a stable cell line where a portion of the intron 1 of the FXN gene containing GAA_{148} was fused to a GFP reporter (Grant et al., 2006). One of the interesting components identified was pentamidine. This drug, already used for patients with HIV, can increase the expression of the GFP reporter, but also the FXN level in cultured patient lymphocytes (Grant et al., 2006). These data confirm the potential of small ligands as therapeutics.

Probably the best ligand to bind GAA secondary structure would be a specific antibody. One group has already tried to produce specific monoclonal antibody against triplexes. Using synthetic oligonucleotides containing GAA repeats, a secondary structure was associated *in vitro* and used to immunize mice (Agazie et al., 1994; Lee et al., 1987). The *in vitro* assays to characterize these antibodies were promising, but the *in vivo* data were generally less reliable (Ohno et al., 2002). Apparently this work was abandoned several years ago and the antibodies are not available anymore (personal communication from Lee). However, such an antibody could play important roles for Friedreich's ataxia patients. Maybe it would have the potential to bind *in vivo* triplexes and restore the frataxin transcription. It could also be used as a diagnostic tool, which can detect the expanded repeats and their ability to form secondary structure, leading to more precise prognosis. Finally, even if the clinical applications are not doable, an antibody specific for triplexes would be a very helpful tool for scientists to detect the genetic factors involved in TNR instability. For these reasons, one of the aim of this thesis has been to develop an antibody specific to GAA secondary structure.

1.6.4. GAA repeats instability

If the way by which GAA repeats cause Friedreich's ataxia is more or less understood, the reason why the repeats expanded in the first place remains unclear. Different models are proposed to explain TNR instability. They involve the interference of GAA secondary structure with replication, transcription or repair mechanisms, respectively. Each of them is discussed below.

*1.6.4.1. GAA secondary structure: YR*R triplex*

"A DNA triplex is formed when pyrimidine or purine bases occupy the major groove of the DNA double helix forming Hoogsteen pairs with purines of the Watson-Crick basepairs" (Fig8B) (Frank-Kamenetskii and Mirkin, 1995). Different forms of triplexes exist. First we can distinguish between triplexes made of two pyrimidine and one purine strands (YR*Y) (Fig8A bottom) and the one made of one pyrimidine and two purine strands (YR*R) (Fig8A top). YR*Y triplexes are more stable under acidic conditions, because they need to be

protonated, whereas YR*R on the other side form at neutral pH, depend on the presence of bivalent metal cations and are stabilized by negative supercoiling (Frank-Kamenetskii and Mirkin, 1995). Second, a triplex can be intermolecular, coming from the interaction of two different molecules, or intramolecular, coming from one single strand folded on itself. Third and last, triplexes can also involve one ssRNA strand and a DNA double helix.

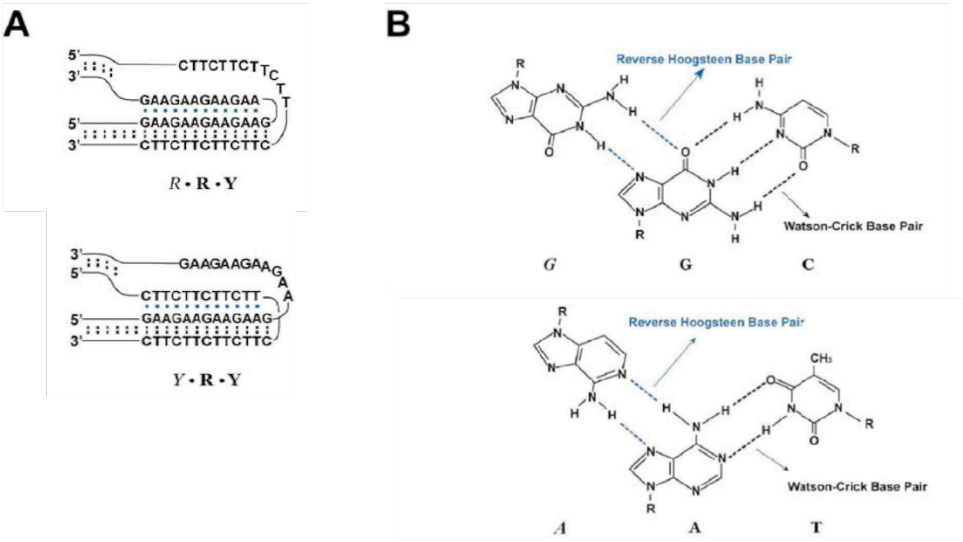


Figure 8. Triplex DNA structure formed by GAA/TTC. A. Two probable H-DNA configurations adopted by GAA repeats. The *RR*Y* triplex is more stable and more versatile than YR*Y. B Base pairing within G.G.C and A.A.T triplexes. The Hoogsteen hydrogen bonds are shown as blue dots and the Watson-Crick base pairings are represented as black dots.

We know since the 1980s that supercoiled plasmids containing GAA repeats can form intramolecular triplexes (Hanvey et al., 1988), whereas GAA intermolecular triplexes were observed only in 1999 (Sakamoto et al., 1999; Vetcher et al., 2002). While running linearised plasmids containing different number of repeats on an agarose gel, the group of Robert Wells identified, when GAA repeats were >50, a second band with slower migration (RB=retarded band). With a combination of different digestions and electron microscopy they showed that RB corresponds to two plasmids connected at the repeats. They then analysed the different characteristics of the junction by assessing the stability of RB. They found that negative supercoiling or a pH=8 stabilizes the structure, whereas a pH=5 or addition of EDTA, which sequesters metal ions, destabilizes it. Moreover, P1 nuclease, which cuts ssDNA, processes the junction with a higher affinity (13.5x) for the TTC strand. Taken together all these data suggest that GAA repeats can form intermolecular triplexes in *E.coli*.

1.6.4.2. DNA replication and GAA repeat stability

Bacteria and yeast

Because DNA replication is considered responsible for microsatellite expansion or contraction by a process known as "polymerase slippage" (for review see Kovtun and McMurray, 2008), it was also suspected to affect TNR length. Studies on bacteria (Pollard et al., 2004) and yeast (Shishkin et al., 2009) confirmed that replication of DNA containing more than ~50 GAA repeats leads to instability. Moreover, the orientation of the GAA tracts according to the origin of replication seems to play a role in the mutation rate: GAA repeats undergo large deletions when located as the lagging strand template (Pollard et al., 2004), Shishkin 2009). For the expansion the frequency of events are similar in both orientations and rare. Considering this information and the correlation between the appearance of the mutations and the formation of secondary structure, a first model was suggested (Fig9) (Pollard et al., 2004). When GAA repeats are on the lagging strand template during replication, they have the risk to be temporarily single stranded, due to the processing of the Okasaki fragment. They could then have the possibility to fold on themselves causing a polymerase slippage and a large deletion. To obtain an expansion, GAA should form a secondary structure when located on the newly synthesised strands, which is less frequent due to the absence of ssDNA steps.

The potential role of DNA replication on repeat instability is also supported by yeast genetic analyses (Shishkin et al., 2009). They showed that Δtof1 and Δcsm3, which are TIM and TIPIN homologues (replication pausing complex), increase GAA expansion of ~6x and ~3x respectively.

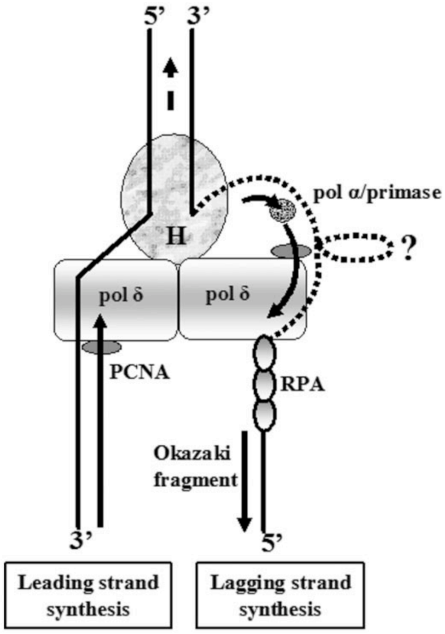


Figure 9. Model depicting the genesis of large contractions during lagging strand synthesis of GAA triplet repeat sequences (in eukaryotic replication). 'H', helicase; 'PCNA', proliferating cell nuclear antigen; RPA, replication protein A. Polδ (large rectangle) replicates the leading strand at the advancing fork. Polα/primase (gray filled circle). (Pollard et al., 2004)

While DNA replication is suspected to mediate the initial GAA expansion, once expanded GAA repeats induce detectable replication fork stalling. In *E. coli,* a slow down of the replication is observed when expanded GAAs are on the lagging strand template (Pollard et al., 2004). In *S. cerevisiae*, 2 studies proved by neutral-neutral 2D-gel analysis that GAA repeats pause the fork in a length- and orientation- dependent manner (Krasilnikova and Mirkin, 2004; Shishkin et al., 2009). However, there is no direct correlation between the orientation of the replication fork pausing and TNR expansion (Shishkin et al., 2009), which may suggest that triplet instability is unrelated to fork pausing and rather mediated by post-replicative events (see point 1.6.4.3)

Mammals

The main studies trying to assess the relationship between DNA replication and TNR instability *in vivo* in mammalian cells are based on replicating plasmids. The following paragraphs will discuss the different results obtained by this model and its positive and negative points.

The plasmid-based system takes in general advantage of the simian virus 40 (SV40) properties. Its genome is condensed on a circular double stranded DNA of 5.2kb (Fig10), which can replicate many times per cell cycle. The explanation to such a short genetic code comes from the ability of SV40 to recruit the host proteins. This means that its duplication mechanism is very similar to mammals. However, one major difference distinguishes the two models: the helicase. If higher eukaryotes use Mcm2-7 complex to unwind the DNA, SV40 needs a protein called large T-antigen (Tag). Besides its essential role in viral replication, Tag can stimulate host cells to enter in S-phase and blocks important tumor suppressors like p53 and pRb (for review see Vilchez and Butel, 2004). These two functions make the analysis of cell cycle and repair proteins unreliable upon SV40 transfection. In summary, SV40 replication *in vivo* is very similar to the mammalian system and requires only the SV40 origin and the large Tag.

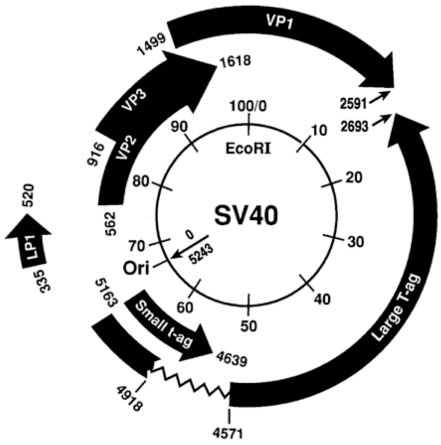


Figure 10. Genetic map of SV40. The circular SV40 DNA genome is represented. The open reading frames that encode viral proteins are indicated. (Vilchez and Butel, 2004)

Based on this information, the studies addressing TNR replication *in vivo* have previously used bacterial plasmids containing the SV40 origin, which they transfect in specific cell lines expressing the Tag (HEK293T, COS1, COS7,…). Despite the ability of these plasmids to make several rounds of replication per S-phase, their amplification is not very efficient and the amount of DNA recovered is in the range of nanograms/cell culture dish. For this reason they need either to be analysed by Southern blot or to be retransformed in bacteria for further

characterization (Cleary et al., 2002). One of the main advantages of this system is that different numbers of repeats can easily be cloned on the vector in order to test the effect of their length, orientation or position according to the origin (Cleary et al., 2002; Ohshima et al., 1998). As in bacteria and yeast, data show that plasmid DNA replication *in vivo* is impaired by long GAA repeats (Ohshima et al., 1998). The position of the repeats in respect to the origin also plays a role, but the causative effect is still unclear (Rindler and Bidichandani, 2011).

1.6.4.3. Post-replicative repair (PRR)

PRR, as the name suggest, is a repair mechanism acting behind the replication fork. It is regulated by Rad6-Rad18 epistasis group and helps the replicating polymerase polε to bypass lesions by two different pathways: the translesion synthesis (error-prone) and the template switch (error free). The translesion synthesis recruites specialized polymerases, which can replace polε and polymerise over the damage. This pathway is error-prone, because the bases introduced by the translesion polymerases are not always correct, causing mispairing and possibly mutation. Despite its mutagenic characteristic this pathway is probably not involved in the large genomic instability observed at TNR.

Template switch "is hypothesized to involve a switch of templates in which the blocked nascent strand uses the undamaged sister chromatid as a temporary replication template" (Branzei 2011). The mechanism by which the switch occurs is still unclear, but two different models have been proposed. The first one suggests the formation of fork reversal, limiting the effect of PRR directly at the fork. The second requires a recombination-like invasion mechanism able to act at the fork or behind the fork. The two models are not necessarily mutually exclusive.

Deletion of the proteins mediating the template switch (Rad5 or Rad6) was shown to reduce TNR expansion in yeast (Daee et al., 2007; Shishkin et al., 2009). Thus, PRR seems to be mutagenic at the repeats.

Different template switch models have been proposed to explain TNR instability. The first one is based on fork reversal (McMurray, 2010; Shishkin et al., 2009) (see Fig11). Aa. TNR blocks the leading strand polymerase (black circle),

whereas the lagging strand can carry on synthesising (green line). Ab. To bypass the repeats, the two nascent strands regress forming a 4 way junctions (reversed fork). Ac. Using the newly synthesised lagging strand as a template, DNA polε copies enough DNA to bypass the block (dashed line). Ad. During the formation of fork reversal or during fork restart, TNR can form a secondary structure (in green) leading to repeat expansion.

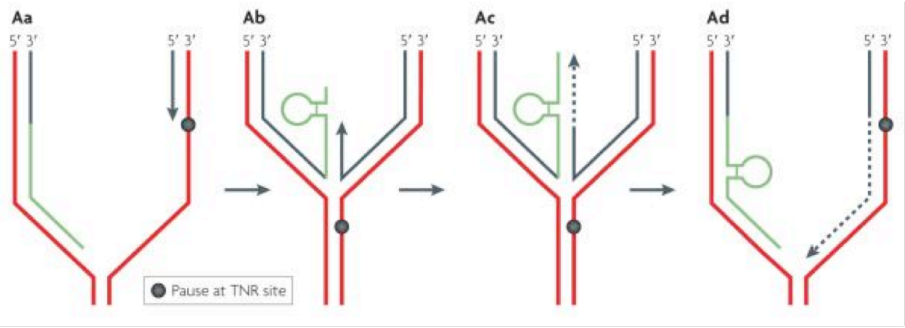


Figure 11. A model for loop formation based on polymerase stalling and restarting within a trinucleotide repeat. (McMurray 2010)

The second model is based on the recombination-like mechanism acting behind the fork (Fig12) (Branzei and Foiani, 2010). During strand invasion and polymerization of the newly synthesized strand the DNA polymerase can undergo misalignement or slippage, which can cause repeat expansion or deletion.

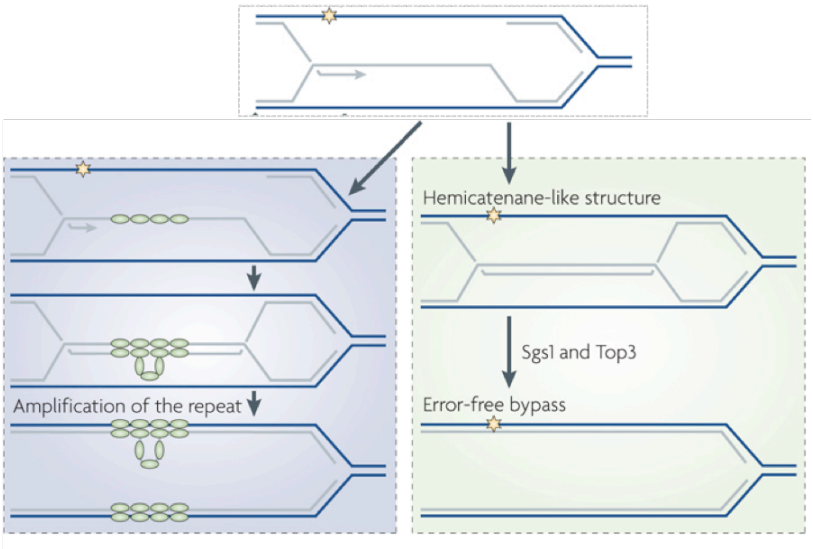


Figure 12. Template switch-mediated damage bypass mechanisms. Template switching occurs through a series of events that, in principle, lead to the error-free bypass of lesions, as indicated in the box on the right. When the region to be copied contains repetitive elements, slippage and misalignment can lead to repeat contraction or amplification. An amplification event is highlighted in the box on the left. (Modified from Branzei and Foiani, 2010)

1.6.4.4. Other mechanisms involved in GAA repeat stability

Transcription

The expansion or contraction of TNR repeats seems to be also affected by different processes like transcription or repair. In Soragni 2008 (Soragni et al., 2008), the authors took advantage of a reporter minigene containing GAA_{560} under the control of an inducible CMV promoter to show *in vivo* in replicating 293T cells that GAA repeats undergo important deletions only when transcription is active. The same results were also observed by Ditch et al. (Ditch et al., 2009) confirming the potential impact of transcription on triplet instability. The authors demonstrated that GAA expansion is independent of cell division rate and demonstrates a role of the transcription mechanism. This same study shows that >88 GAA repeats inserted in HEK293 genome are unstable in a length and orientation manner. They also found that cells containing GAA_{352} mainly undergo expansion than deletion, which is different from what was observed before. This discrepancy can be due to the larger number of repeats used in this assay (352) compared to the one used in bacteria (~80), yeast (100) or SV40-based (115) experiments.

TC-NER

When a lesion stalls the RNA polymerase II (RNApolII), the damage is repaired by a mechanism called transcription coupled nucleotide excision repair (TC-NER). TC-NER is a subpathway of the NER, which removes lesions at transcribed strands only. It is activated by a specific sensor protein travelling with the RNApolII. As NER it repairs by excising, removing and restoring the damaged DNA by polymerization (for review see Fousteri and Mullenders, 2008).

Recent studies have associated the transcription repair mechanism to TNR instability (Lin and Wilson, 2007). Downregulation by siRNA of factors involved in TC-NER in human cells reduces CAG repeats contraction. These data suggest that the normal activity of TC-NER is involved in the instability mechanism.

MMR

The mismatch-repair (MMR) mechanism was also identified as a factor of genomic instability in yeast. Deletion of MMR proteins reduces the rate of large deletions and chromosomal rearrangements at the repeats, while it increases the number of small deletions (Kim et al., 2008). As for TC-NER, these results show that MMR is mutagenic for TNR. However, another yeast study observed an increased expansion rate in Δmsh2 strains (Shishkin et al., 2009). Taken together these data suggest that GAA repeat expansion mechanism can be resolved by MMR, whereas the deletions are favoured by its recruitment (for review see McMurray, 2010).

1.7. Aims of the thesis

The mechanism leading to TNR instability remains unclear. A large number of pathways seem to be involved in repeat expansion or deletion, but none of the models suggested so far can recapitulate the whole picture. Moreover, the differences between the TNR sequences, secondary structures and the organisms add complexity to the published research. However, understanding the molecular mechanism and the genetic factors involved in TNR instability is of crucial clinical relevance. Many patients suffer of rare neurodegenerative diseases caused by triplet expansion, for which no treatment is available. Discovering the molecular determinants of GAA expansion would allow the development of new treatments to cure or at least slow down the disease.

In the first part of this thesis, I aimed to characterize further the replication of GAA repeats *in vivo* in mammalian cells. Using a combination of neutral-neutral 2D-gel and electron microscopy we analysed replication intermediates, which are GAA length- and orientation-dependent.

The second part describes instead the establishment and the characterization of a mouse monoclonal antibody specific for GAA secondary structure. This work was made in collaboration with the group of Prof. Toshio Mori in Japan. This antibody was produced to detect GAA repeat instability *in vitro* or *in vivo* in a direct way, without requiring PCR or sequencing. Such a tool could in principle allow a fast screening and identification of the factors involved in TNR stability.

2. Results I

2.1. A new system to study RIs in mammalian cells

The study of replication intermediates (RIs) in mammalian cells is limited by the undefined location of their origins of replication and by the lack of well-characterized multi-copy loci in their genome. To overcome these limits, we took advantage of SV40 plasmids, which can replicate at high copy number in mammals and use one specific origin. The SV40 genome was ligated to the pGEM plasmid in order to facilitate further cloning, like the insertion of different number of GAA repeats. The presence on the plasmid of the only viral gene required for replication - Large T-antigen (Tag) - dramatically increases its replication efficiency and allows the recovery of significant amounts of RIs from different cell lines, like 293T or U2OS. 40h after transfection of a single 15-cm cell culture dish, micrograms of plasmid DNA containing replication intermediates can be recovered and processed by 2D-gel or EM.

2.2. GAA repeats induce the formation of specific intermediates

GAA repeats were shown to affect DNA replication in bacteria and yeast. In order to know if similar effects can be observed in mammals, GAA repeats of different sizes (33, 66, 90) were cloned in the SV40 plasmid in the two possible orientations (GAA as lagging strand template or TTC as lagging strand template). 40h after transfection in 293T cells, the RIs were recovered and processed for neutral-neutral 2D-gel (Fig13).

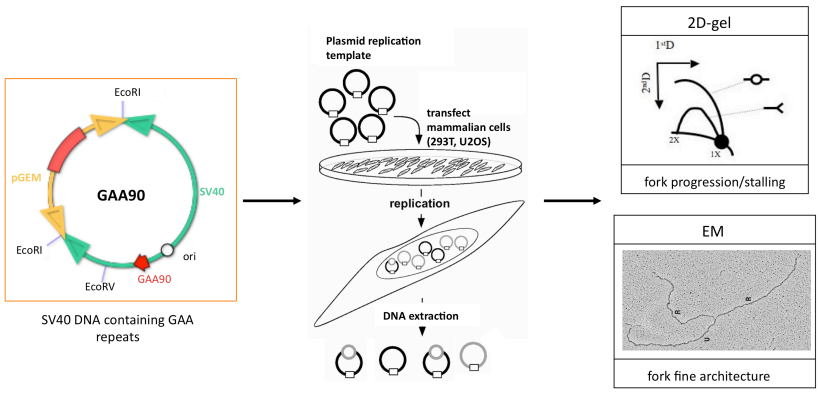


Figure 13. SV40 plasmids containing different numbers of GAA repeats are transfected into 293T cells. After several rounds of replication the plasmids are extracted and the RIs are analysed by 2D-gels and electron microscopy.

The plasmids were digested by EcoRI giving rise to two fragments: one of about 5kb containing the origin of replication and the repeats, and a second one of 3kb corresponding to the area of the fork fusion (Fig14). These two different fragments of the plasmid will be hereafter referred as "initiation zone" and "termination zone" respectively. DpnI restriction enzyme, which is specific for methylated DNA, was also added to the digestion mix in order to degrade the non-replicated plasmids. RIs were then analysed by neutral-neutral 2D-gel.

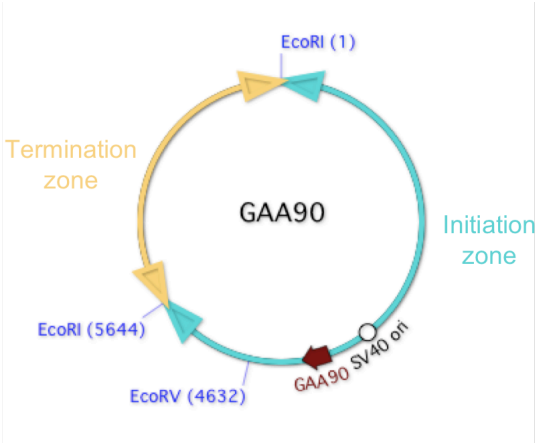


Figure 14. Restriction map of GAA90 plasmid. In light blue is the initiation zone containing the SV40 origin of replication and the GAA repeats. In yellow is the termination zone.

The initiation zone panel (fig15) shows RIs of the control plasmid containing no repeats, the one containing GAA33, 66, 90 repeats and TTC33, 66, 90. As expected (see Materials and Methods), the pattern of the control is composed of a bubble arc, late Ys and a 2n-spike (X molecules). The 2D-gel signal changes with increasing number of GAA repeats: for GAA n≥66, the 2n-spike disappears, and a defined X spot is instead detectable towards the tip of the spike (XS); moreover, a new intermediate is detected under the inflection point of the Y arc (YS); and another prominent one appears above the monomer spot (AMS) (Fig15).

In plasmids with TTC n≥66, the 2n-spike undergoes the same pattern described for GAA repeats. However, in this orientation, the X spot seems to be connected to the inflection point of the Y arc, which is barely visible in the GAA pattern. Moreover, instead of a YS formation, we observe signal accumulation (pausing) at a specific position on the Y-arc (Fig15 pink arrow) and several discrete signals under the Y arc inflection point (SYS). The signal above the monomer spot

increases with TTC repeats, but, differently from the GAA orientation, it is spread over several minor signals (Fig15).

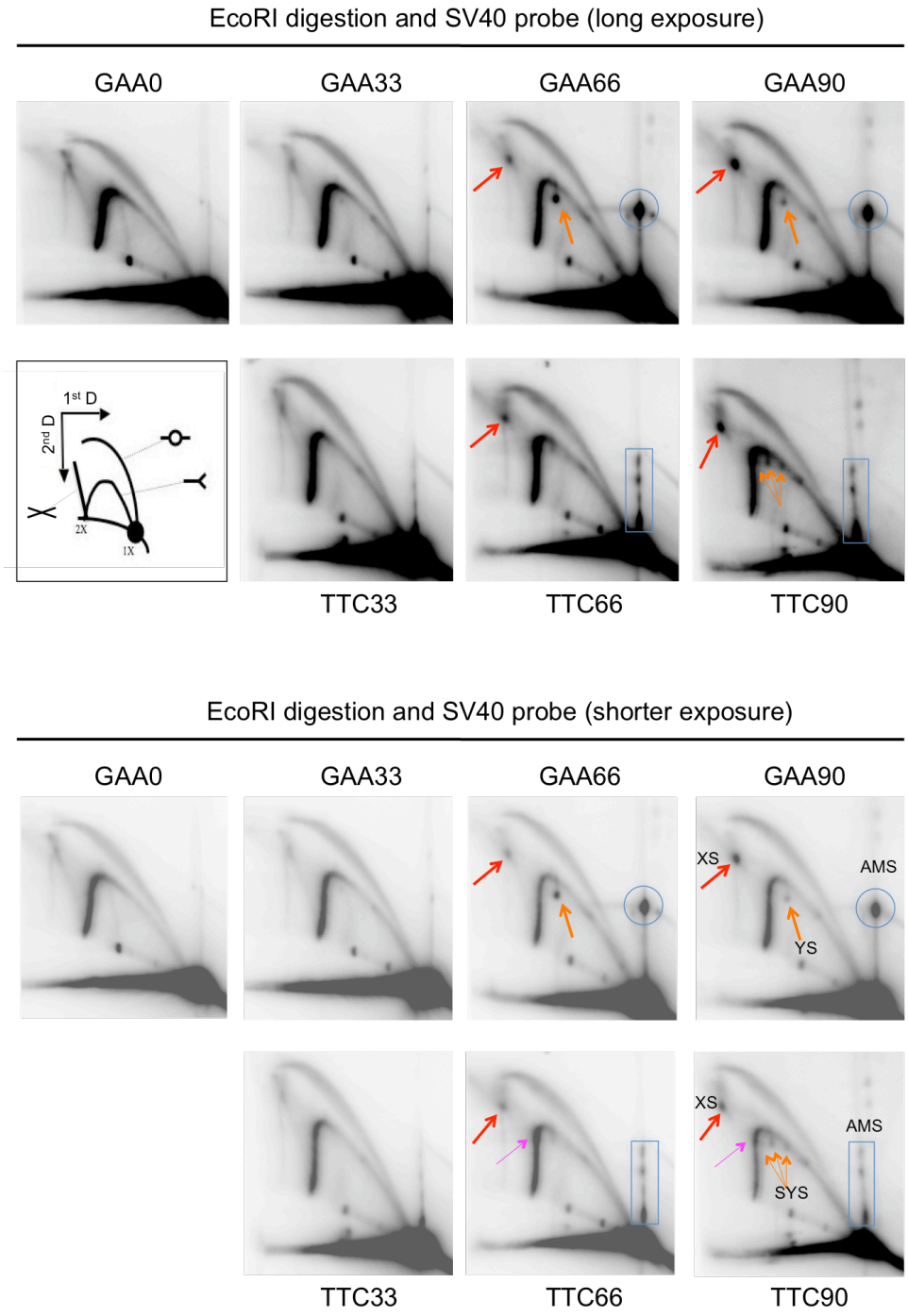


Figure 15. Neutral-neutral 2D-gel analysis of the initiation zone of plasmids containing GAA or TTC repeats as template for lagging strand synthesis. Plasmids were digested by EcoRI, processed by 2D-gel and probed with SV40 DNA. Bottom left part of the top panel: 2D-gel representation. Intermediates specific to GAA/TTC repeats are indicated.

To assess whether the altered RI pattern observed above was indeed related to specific structures arising at the repetitive sequences, plasmids were digested with SacI and EcoNI and run by 2D-gel (Fig16). As expected, the new digestion leads to the formation of a Y-arc and to the absence of a bubble arc, confirming that replication initiation strictly happens at the SV40 origin. The X spot position of GAA90 and TTC90 moved along the 2n-spike according to the different distance of the repeats from the restriction sites (see Materials and methods). The spots formerly migrating under the inflection point of the Y-arc are now visible as pausing signals along the Y-arc (early RIs) (Fig16). This pattern modification will be discussed later (see paragraph 2.7.1).

Taken together these data show that GAA repeats induce the formation of specific intermediates in a length- and orientation-dependent manner.

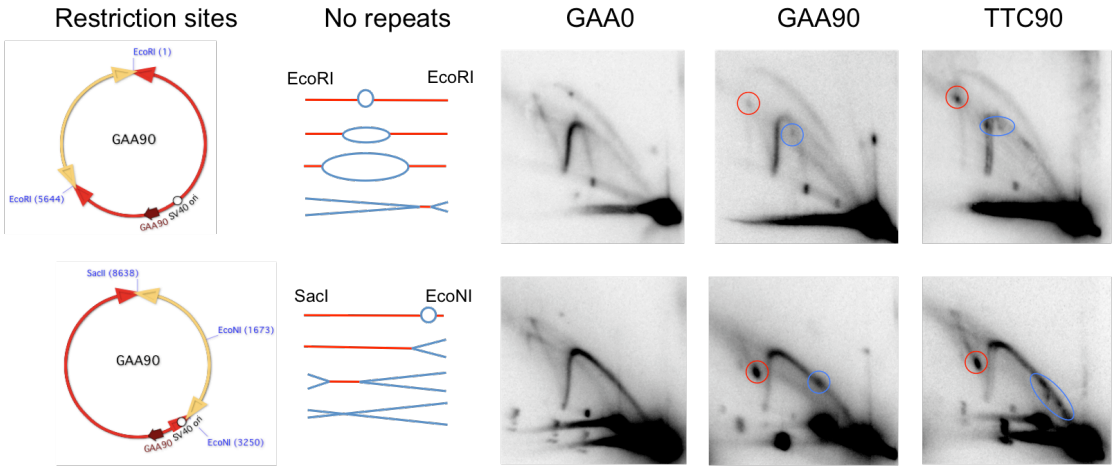


Figure 16. 2D-gel analysis of the control, GAA90 and TTC90 plasmids using different restriction sites. In the upper panel the plasmids are digested by EcoRI and probed with SV40 DNA. In the lower panel the same plasmid DNA preps are digested by SacII-EcoNI and probed with pGEM DNA (see Fig14). The fragments identified by the probes are indicated in red in the restriction maps and the schemes of the intermediates expected for the control plasmid are drawn next to them.

To confirm that these results are not specific to 293T cells, we repeated the experiment using U2OS cells. The obtained 2D-gel patterns look very similar to those described for 293T, suggesting that the effects of GAA/TTC repeats on plasmid replication are general and not cell line-dependent (Fig17).

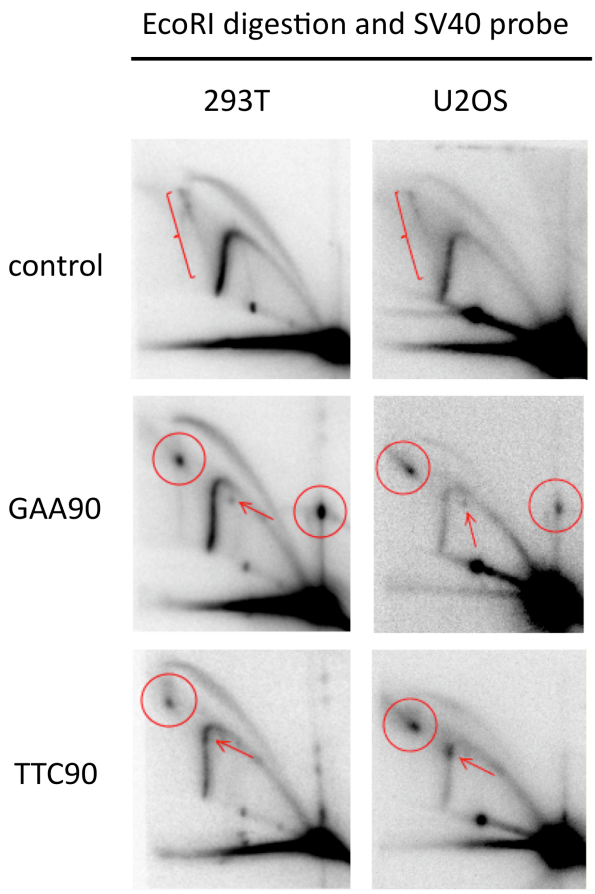


Figure 17. Neutral-neutral 2D-gel analysis of plasmids containing GAA repeats in different cell lines. DNA was digested by EcoRI and probed with SV40 DNA (initiation zone). GAA-dependent changes of 2D-gel pattern are highlighted in red.

2.3. GAA90 and TTC90 repeats induce a transient fork pausing

The 2D-gel analysis revealed unusual RI patterns associated with expanded (>33) GAA repeats. In order to gain insight into the effect of the repeats on replication, we directly visualized RIs from GAA90 and TTC90 plasmids by a specialized electron microscopy approach (EM; see Materials and Methods) (Fig18).

Plasmids extracted from 293T cells were linearised by EcoRV (see Fig14), enriched for RIs on BND cellulose and processed for EM (BAC spreading and platinum shadowing). Each replicating molecule was measured and represented graphically according to its replication status: blue for the replicated part and red for the unreplicated part of the plasmid (see Fig18A). Supported by the 2D gel analysis in Fig15 - which shows replication initiation to be limited to the SV40 origin - I arbitrarily oriented all molecules by having the origin always on the same side (on the left). All intermediates were then aligned from the least to the most replicated, creating from static pictures a dynamic replication map (Fig18A).

With this technique it is possible to compare the kinetic of fork progression on different plasmids. The control panel shows a regular progression of the two forks, which terminate their replication in the region opposite to the origin (Fig18A). In the presence of GAA90 we observe an asymmetry in fork progression, the left one being initially slowed down in the region of the repeats (Fig18B), as shown by the fact that it crosses the EcoRV site at a later replication stage. However, fork fusion in this plasmid still takes place in the same region as for the control plasmid. This result suggests that fork pausing at the GAA repeats is only transient and/or that the second fork is also transiently affected in its progression.

In the case of TTC90 (Fig18C) the effect observed with GAA90 is somewhat exacerbated. A larger fraction of replicating molecules have a fork paused at the repeats and a few plasmids show fork fusion outside of the termination zone.

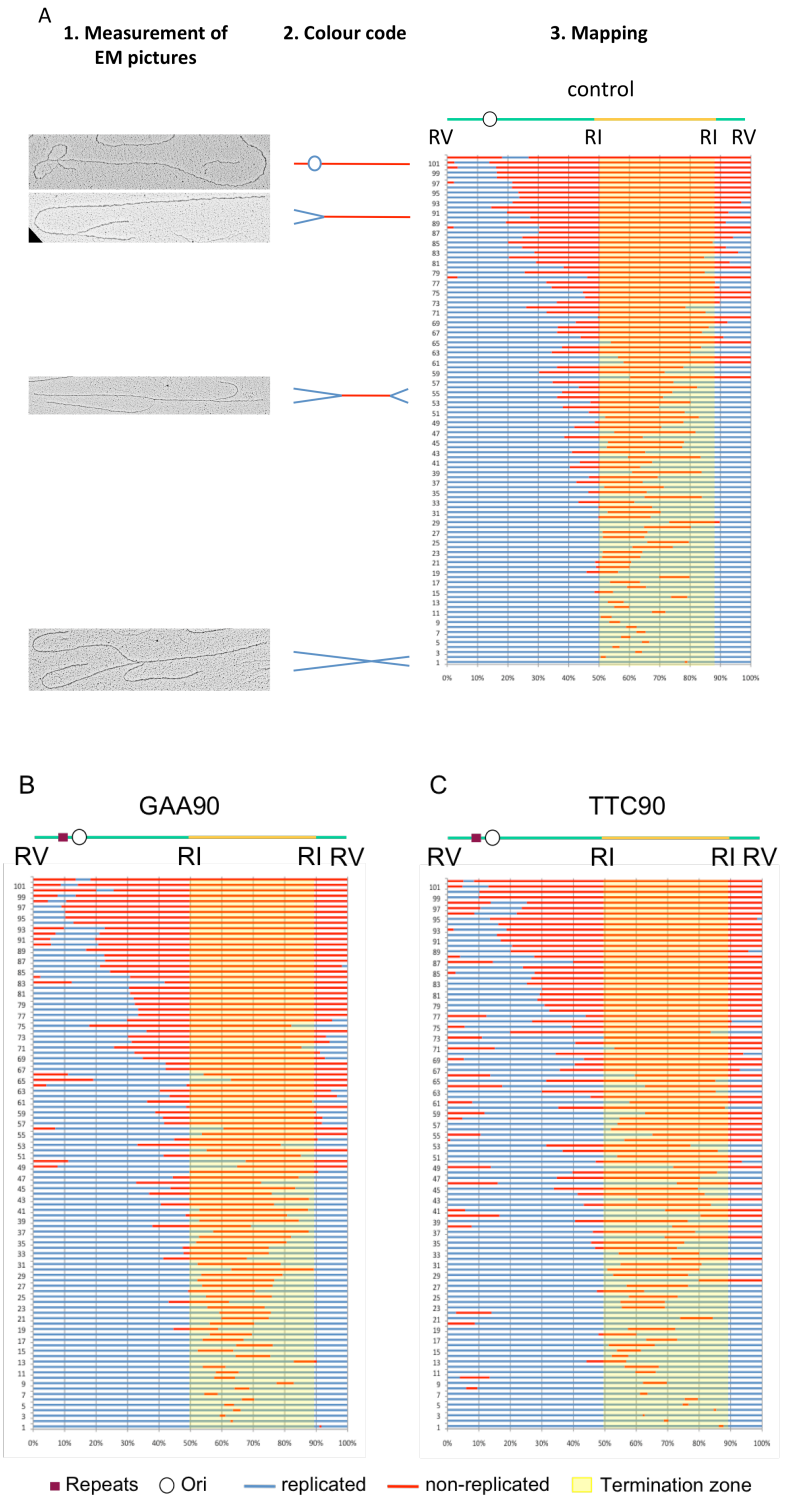


Figure 18. Dynamic replication map of GAA0, GAA90 and TTC90 plasmids in 293T cells. A. Description of the mapping: 1) Replicated and unreplicated zones of replication intermediates are identified and measured in the EM pics. 2) Based on these measurements RIs are converted into schematic representations indicating in blue the replicated zones and in red the not replicated 3) All the schematic representations are oriented with the origin on the left side and then aligned from the least to the most replicated. B. GAA90 replication map. C. TTC90 replication map. (Each map is built with ca. 100 molecules)

The fact that most of the plasmids, either with GAA or TTC repeats, finish their replication in the expected region is also confirmed by 2D-gel. In the termination zone (Fig19) all 2D-gels show the same composite pattern, a Y-arc flanked by a "cone signal", indicating that a second fork enters the fragment while the first is still replicating it - a diagnostic pattern of replication termination.

These data prove that GAA and TTC repeats affect only transiently the passage of the replication fork, as termination of plasmid replication is still mostly achieved opposite to the origin and thus far from the expanded GAA repeats. However, TTC seems to have a stronger effect than GAA, as it may occasionally arrest the fork until fork fusion.

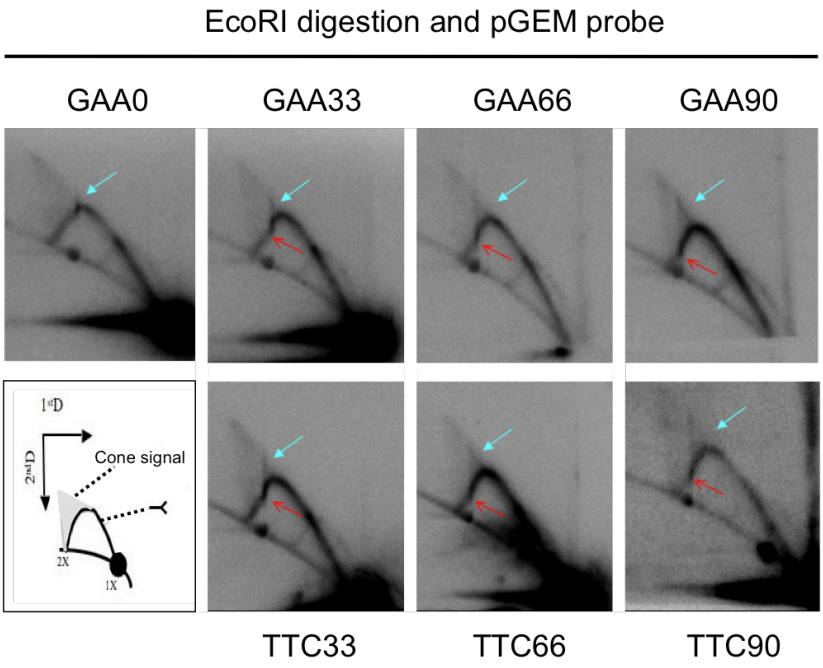


Figure 19. Neutral-neutral 2D-gel analysis of the termination zone of plasmids containing different number and orientation of GAA repeats.

2.4. Direct visualization of RIs isolated by 2D-gel

The study of a total RI population by EM, as in Fig18, allows the observation of the general effect of the repeats on plasmid replication. However, rare events or transient intermediates are difficult to identify and analyse, as they are under-represented in the whole population.

In order to overcome this limit, we established an experimental system by which 2D-gel intermediates can be directly visualized by ethidium bromide staining, extracted from the gel after 2nd dimension, electro-eluted and finally observed by EM (Fig20). To prove the reliability of the system, the monomer spot and part of the Y-arc were extracted and analysed. The results showed that the monomer spot sample was uniquely composed of linear molecules, whereas the Y-arc sample contained a vast majority of Y-shaped molecules (data not shown).

These controls confirm the efficiency of this technique in recovering sufficient DNA for EM analysis and in maintaining the integrity of the RIs. We thus used this approach to investigate the molecular architecture of the intermediates specifically induced by expanded GAA repeats.

Figure 20. Extraction procedure of 2D-gel intermediate and work-flow for EM visualization.

2.5. Characterisation of the X molecules

2.5.1. Molecules migrating as a 2n-spike are 2 fully replicated plasmids connected by a homology independent junction

According to the known principles of DNA 2D-gel electrophoresis (Brewer and Fangman, 1987) molecules migrating as a 2n-spike must be X-shaped molecules, composed of two interconnected linear fragments. Upon certain DNA extraction procedures, such intermediates are consistently detected - besides standard RIs - during unperturbed DNA replication (see Introduction). Different mechanisms have been speculated to induce their formation, such as recombination, termination, post-replicative repair. Using 2D-gel extraction we expected to get further insights into the molecular architecture of these intermediates.

EM analysis of the 2n-spike reveals indeed the presence of X-shaped molecules (Fig21A). Measurement of the 4 arms of these intermediates shows that these Xs are composed of 2 fully replicated plasmids (arm1+arm2 = arm3+arm4 = ~5.2kb) connected together (Fig21B). In order to investigate the determinants of this connection, we determined whether the junction was symmetrically positioned within the molecules. We compared the size of the two longest arms of the sister plasmids (1 and 3), calculating for each molecule the ratio between the longest and the second longest arms. Symmetric X-molecules are expected to result in a ratio of 1.0, whereas a ratio >1.0 would indicate asymmetric junctions (Fig21C). The data show that 50% of the values are between 1.028 and 1.185, which clearly overcomes the 5% error generally observed for this kind of measures (Lopes, 2009). This evidence indicates that a relevant fraction of the molecules is asymmetric (see also results described in 2.5.2), suggesting that the two linear molecules are connected by a sequence-independent junction, thus excluding homology-mediated structures like Holliday junction, reversed forks or termination intermediates. Accordingly, plotting the arm size in the molecule population indicates that the junction connecting the two monomers can form anywhere along the plasmid (Fig21D).

Careful inspection of those molecules where the junction was more visually accessible (Fig21A) showed that only one ssDNA of the first plasmid contacts the double helix of the second plasmid.

Taken together these results exclude the possibility that the 2n-spike consists of recombination or termination intermediates, and rather suggests the presence of homology-independent post-replicative junctions.

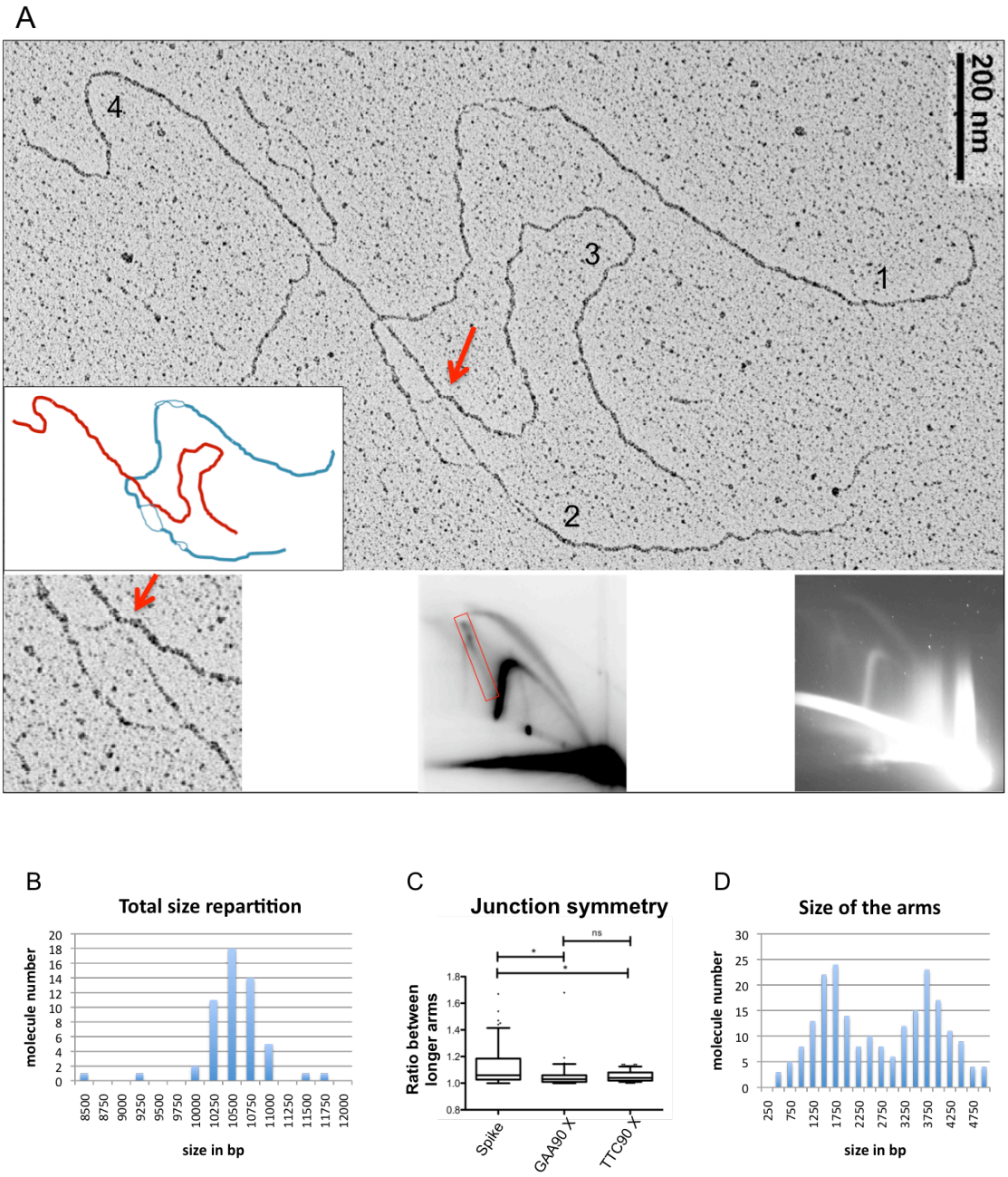


Figure 21. EM analysis of the molecules migrating as a 2n-spike in 2D-gels of control plasmids. A. Electron micrograph of an X-shaped molecule (magnification 46kx). B. Size repartition of the X molecules. C. The symmetry of the X molecule was determined by measuring the ratio between the two longest arms. (Ratio=1.0, the X is symmetric; ratio>1.0, X is asymmetric). (⋆= P<0.05; ns=not significant) D. Each individual arm of the X molecules was measured and plotted.

Accordingly, during the total population EM analysis of control plasmids we observed, in a few rare cases, the presence of a sister chromatid junction behind an active fork (Fig22). The very low number of such molecules does not allow any statistical analysis, but the detection of these intermediates supports the interpretation that post-replicative junctions can form and travel behind replication forks, in agreement with previous reports (Lopes et al., 2003).

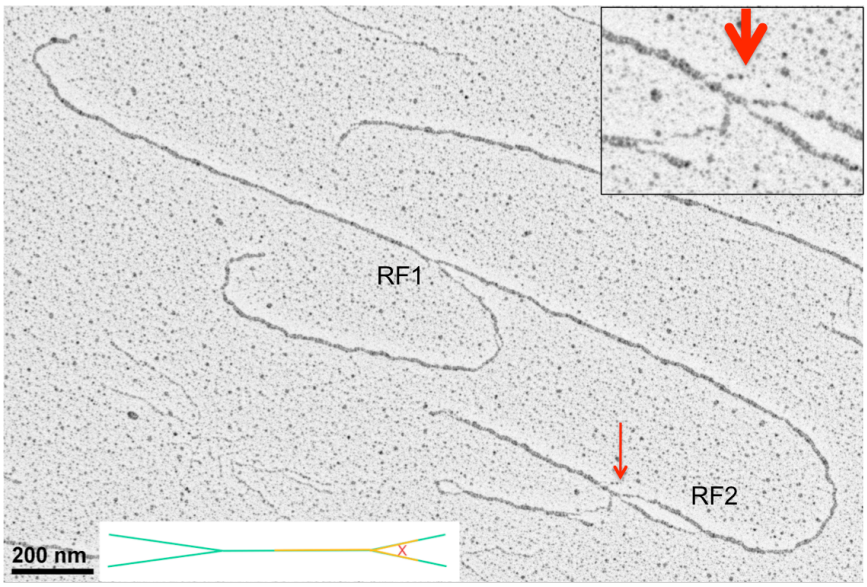


Figure 22. Electron micrograph (46kx) of a replication intermediate with a post-replicative junction (red arrow). RF=replication fork.

2.5.2. XS intermediate is a post-replicative junction located at the GAA repeats

The 2D-gel analysis described in fig15 shows that GAA repeats lead to the disappearance of the 2n-spike and the formation of a specific X-type intermediate (XS). This signal could be in principle attributed either to fork stalling and termination at a specific site or to a persistent post-replicative junction once the forks have completed replication of the fragment.

In order to distinguish between these two hypotheses, we extracted XS from a 2D-gel on GAA90-plasmid and processed it by EM (native spreading) (Fig23A). As for the 2n-spike, size measurements show that these molecules are composed of two joint, fully replicated plasmids (Figs23 B and C). However, differently from the 2n-spike, the interaction between the two sister DNAs is always located at the same position (1/3:2/3 of the plasmid), which corresponds to the site of

the repeats. As already done above, the symmetry of the molecules was tested by calculating the ratio of the two longest arms (Fig21C). 50% of the values lie between 1.01 and 1.06, which is significantly more symmetric than the spike molecules (p=0.0112). Thus, differently from the 2n-spike, sister chromatid junctions on GAA90 plasmids tend to involve homologous sequences, i.e. the repetitive tract itself. Most of the junctions display the interaction of one ssDNA with the double helix of the second plasmid. However, a low fraction of the junctions appear more unwound and allow to notice that the junction can involve just one single DNA strand of each sister plasmid. Moreover, the zone of interaction between the two ssDNAs suggests a possible annealing between the two strands involved in the junction (Fig23A).

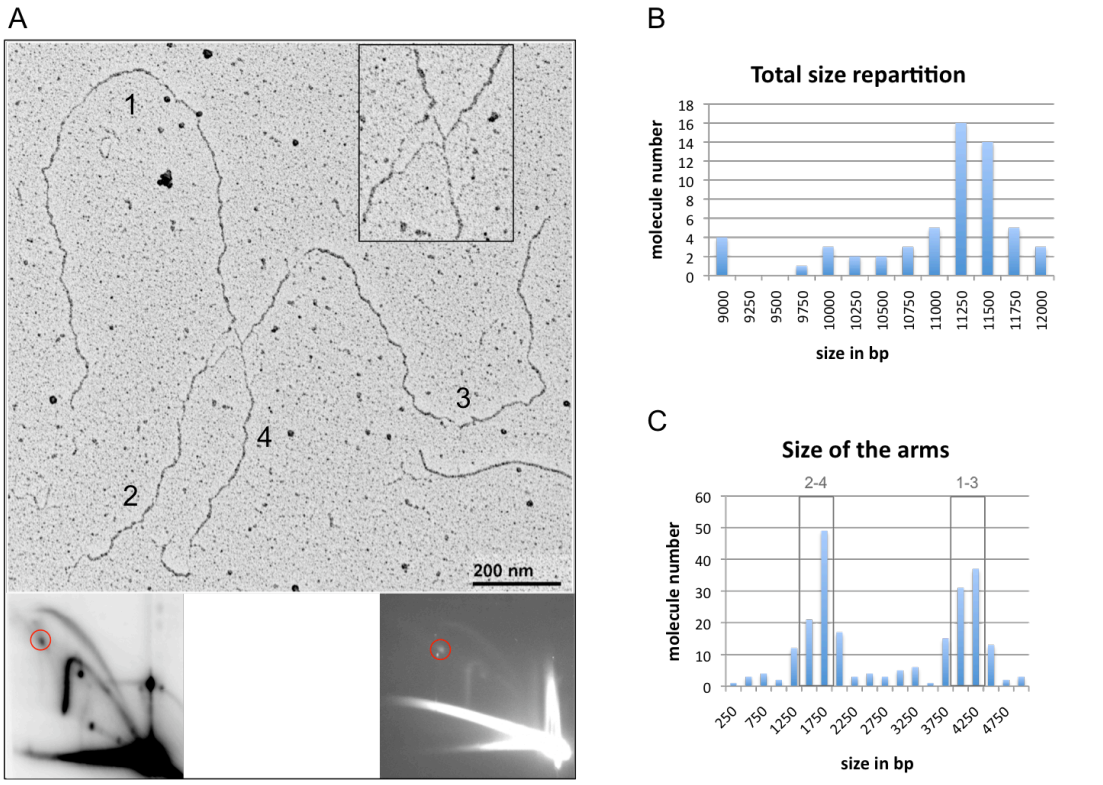


Figure 23. EM analysis of the molecules migrating as a X-spot in GAA90 plasmids. A. Representative electron micrograph of a molecule migrating in this zone (magnification 46kx). Next to it, typical 2D-gel pattern of TTC90 plasmids by Southern blot and EtBr staining of a preparative 2D-gel. The gel portion used for EM analysis is marked by the red label. B. Size repartition of the X molecules. C. Each individual arm of the X molecules was measured and plotted. The arms corresponding to the different sizes are indicated in grey.

2.5.3. TTC90 repeats induce the formation of XS molecules, similar to those observed in GAA90.

XS molecules formed in plasmids containing expanded TTC repeats look similar to the ones observed in GAA90-plasmids (Fig24A). They are also composed of 2 fully replicated plasmids joint at the level of the repeats (Figs24 B and C). The symmetry measurement indicates that 50% of the TTC90-XS population is between 1.02 and 1.08, which is again significantly more symmetric than what observed for the spike molecules (p=0.014), but indistinguishable from GAA90-XS (p=0.2913) (Fig21C).

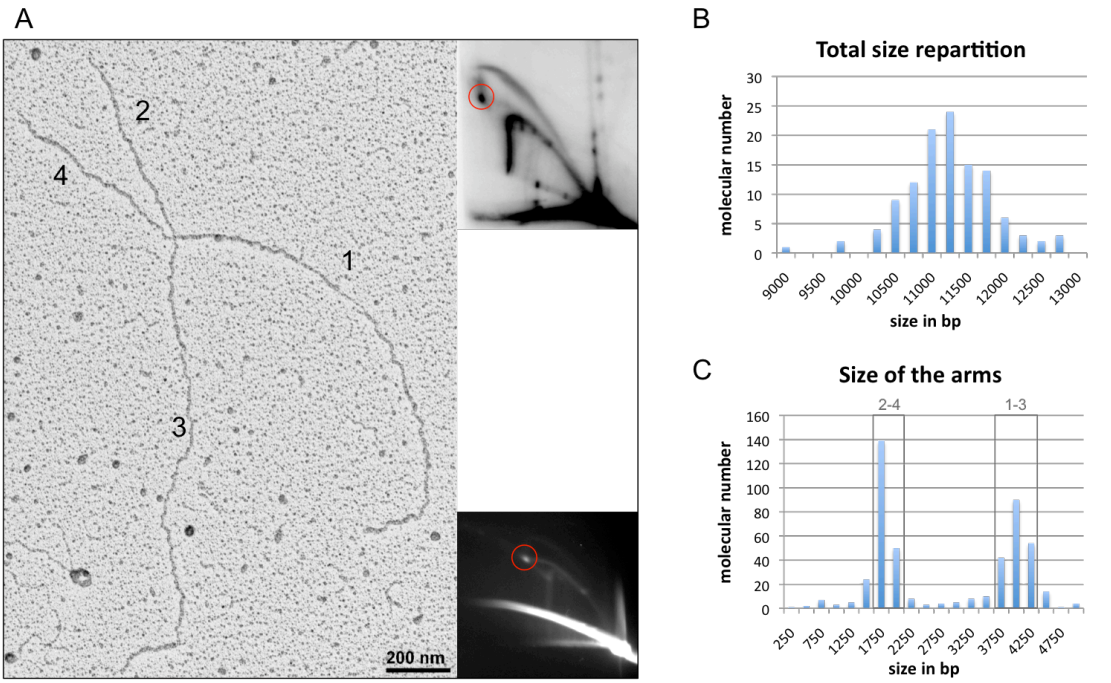


Figure 24. EM analysis of the molecules migrating as a X-spot in TTC90 plasmids. A. Representative electron micrograph of a molecule migrating in this zone (magnification 46kx). Next to it, typical 2D-gel pattern of TTC90 plasmids by Southern blot and EtBr staining of a preparative 2D-gel. The gel portion used for EM analysis is marked by the red label. B. Size repartition of the X molecules. C. Each individual arm of the molecules were measured and plotted. The arms corresponding to the different sizes are indicated in grey.

2.5.4. XS intermediate also forms in bacteria

The presence of symmetric X molecules connected at GAA repeats was already observed in bacteria (Sakamoto et al., 1999). Plasmids containing >60 GAA repeats were shown to form a secondary structure retarded (RB) in agarose gel. Further characterizations by EM proved that this RB was made of two plasmids joined at the repeats. However, the EM approach (Kleinschmidt method) used in that work does not provide sufficient resolution to observe the molecular architecture of the junction. The DNA is covered with cytochrom C protein during sample preparation: while this increases DNA thickness and assists detection, it also impairs the distinction between ss- and ds-DNA.

In order to get further insights into these bacterial intermediates and see if they have structural similarities to the one observed in mammals, we repeated the experiment described in Sakamoto 1999, using our own EM protocol (Fig25). Our pictures show that the junction between the two bacterial plasmids involves one DNA strand of a molecule and the double helix of the second one, similarly to what observed in human cells (2n-spike and XS, Figs 21, 23, 24). As shown for XS, the junction is detected at a specific position along the DNA fragments, consistent with the location of the expanded GAA repeats.

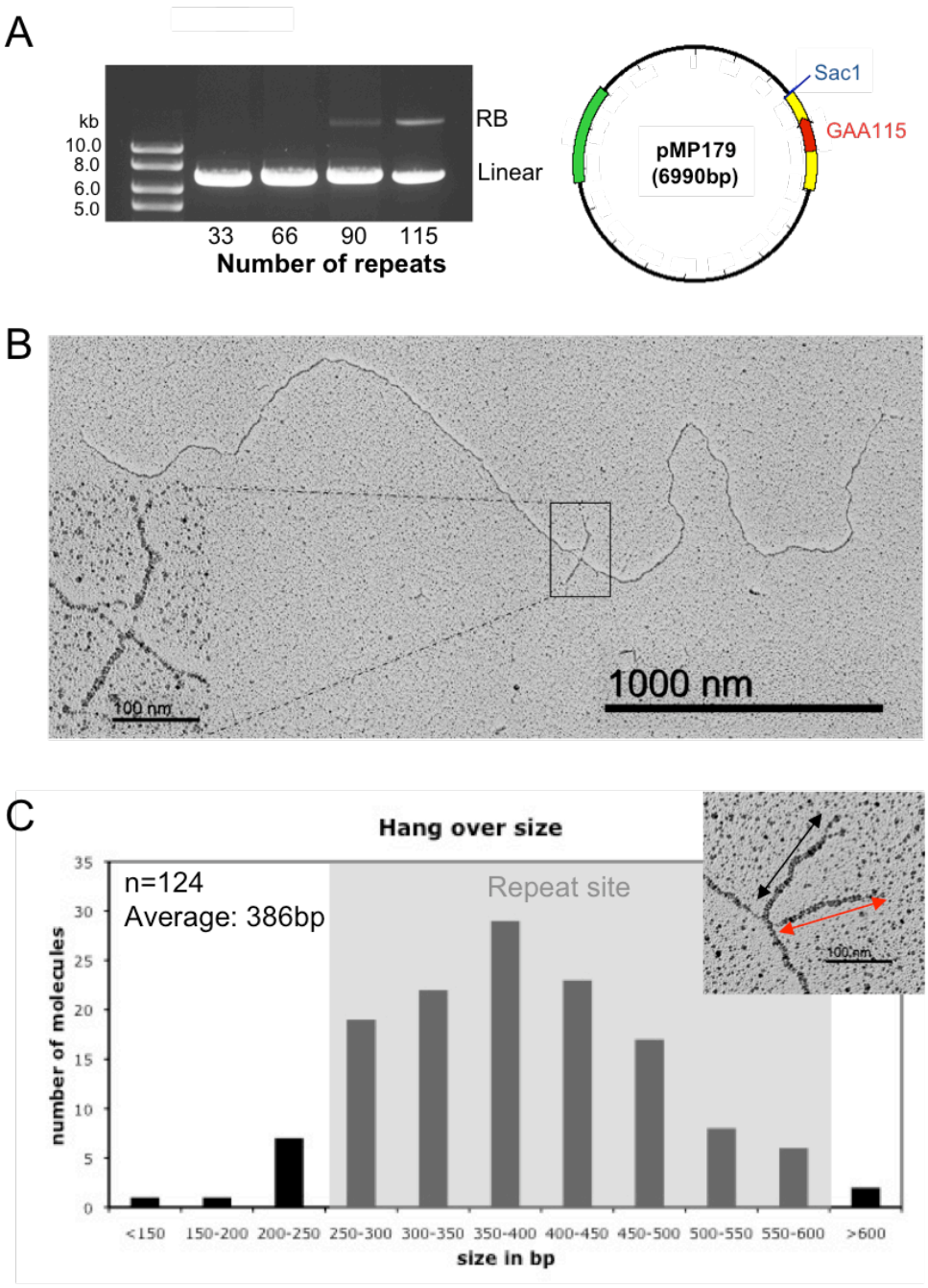


Figure 25. Plasmids containing >60 GAA repeats form a secondary structure in *E.coli*. A. Agarose gel of plasmids containing GAA_n repeats, linearised by SacI, and GAA_{115}-plasmid map (RB= retarded band). B. EM visualisation of GAA_{115}-RB. (Magnification : 46kx for the larger picture and 180kx for the smaller insert). C. The position of the junction along the plasmid has been determined after linearization with SacI and length determination of the short extremities emanating from it. We found that it coincides exactly with the plasmid location of the repetitive GAA sequences.

2.6. X spot formation in TTC90 plasmids

The high quality of the preparative 2D-gels (Figs 21, 23, 24) allowed the extraction of additional signals detected by EtBr staining, such as those connecting the X spot to different positions of the Y-arc in the TTC90 plasmids. As these signals connect two major intermediates detectable in these 2D-gels, we reckoned that the analysis of these molecules could provide insight into the formation of the XS molecules from standard replication forks (Ys). Two different arcs were visible by EtBR: 1) from the pausing signal on the Y to XS; 2) from the inflection point of the Y to XS.

2.6.1. From the pausing signal on the Y to XS

Observation of these intermediates at the EM revealed the presence of two types of molecules (Figs26 A and B). The first population is represented by 3 way junctions, corresponding to large forks of a defined size (Ys). The second population consists of the same large Ys showing 1 supplemental arm (Fig26A arm4). Contour lengths measurements (Figs26 C and D) show that the replicated duplexes measure ~3.9kb (arms 2-3) and the parental ~1.7kb (arm1), which corresponds to a replication fork stalled at the repeats. The 4th arm emerging from the fork was heterogeneous in length and up to ~1.3kb long (Fig26E), representing most likely the regressed arm of a reversed fork (see Introduction). Theoretically, if the reversal happened after the fork had passed through the repeats, the regressed arm could reach up to ~1.7kb length, thereby assuming the molecular structure observed in the X spot (Figs 23 and 24). These last molecules were probably extracted with the XS intermediates explaining why they are missing here.

These data show that fork reversal can happen at expanded GAA repeats and that reversed forks may represent precursors of the molecules accumulating in the X spot.

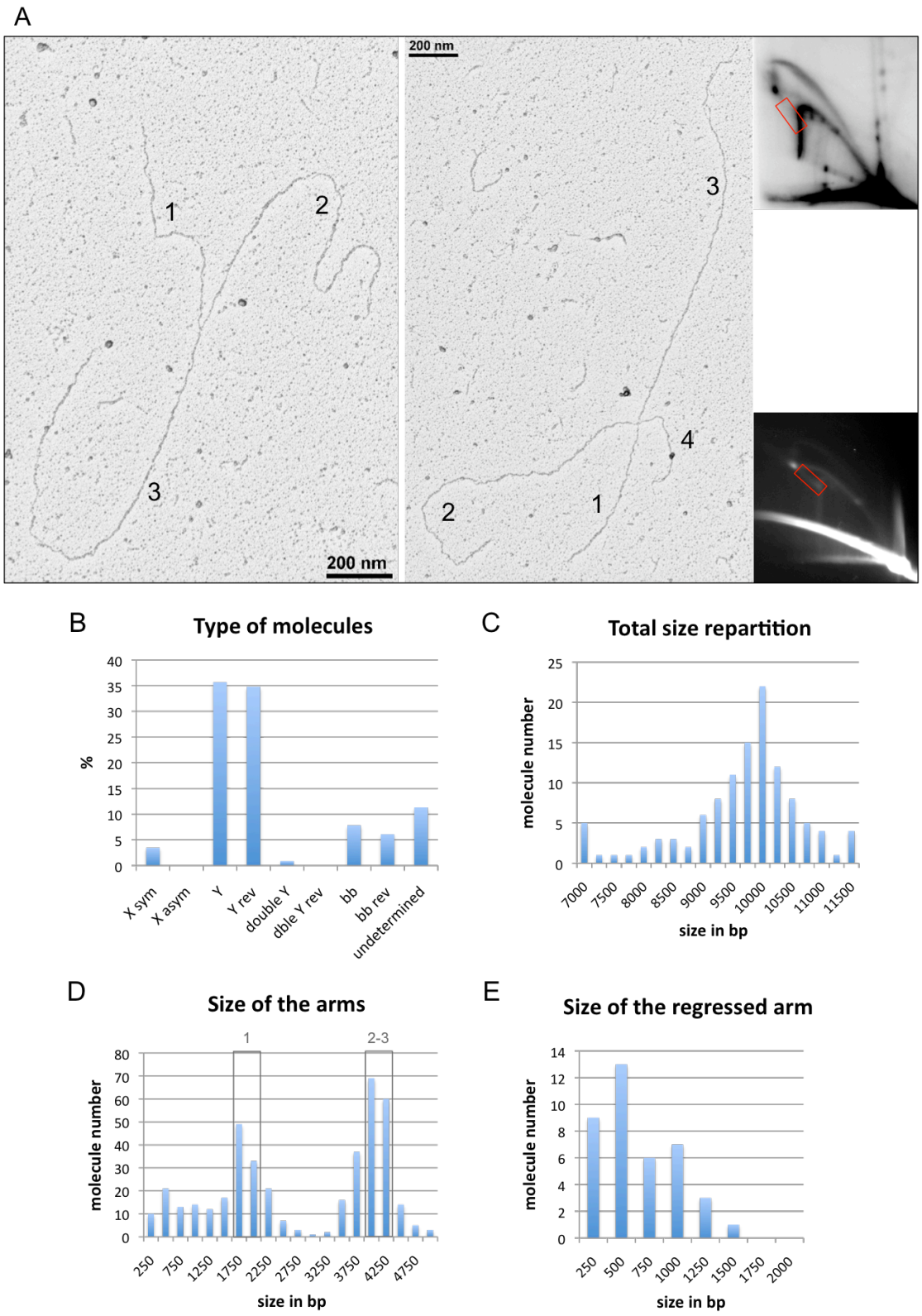


Figure 26. EM analysis of the molecules migrating between the pausing on the Y-arc and the X-spot in TTC90 plasmids. A. Representative electron micrograph of a molecule migrating in this zone (magnification 46kx). Next to it, typical 2D-gel pattern of TTC90 plasmids by Southern blot and EtBr staining of a preparative 2D-gel. The gel portion used for EM analysis is marked by the red label. B. Type of the molecules migrating in this zone. C. Size repartition of the molecules. D. Each individual arm of the molecules was measured and blotted. The arms corresponding to the different sizes are indicated in grey. E. Size of the regressed arm.

2.6.2. From the inflection point of the Y to XS

The intermediates migrating from the inflection point of the Y arc to XS appear as X molecules, with the intersection of the 4 asymmetric arms located at the repeats.

The majority of the molecules have one arm of ~1.7kb and one of ~3.9kb, corresponding to the size of the SV40 genome with TTC90 insert (5.6kb) (Figs27 B and C). The next two arms have variable sizes, of ~1.3-2.0kb (Fig27D) for the shorter and ~2-2.7kb for the longer (Fig27E).

The presence of 4 different arms excludes the possibility that these molecules represent standard replication or recombination intermediates. We suspect them to result from breakage of a bubble with the other fork reversed (Figs27 F and G). In this scenario the arms of ~1.3-2.0k would most likely represent the regressed arm, while the one of ~2-2.7kb would represent the broken arm of the bubble (Figs27 D and E).

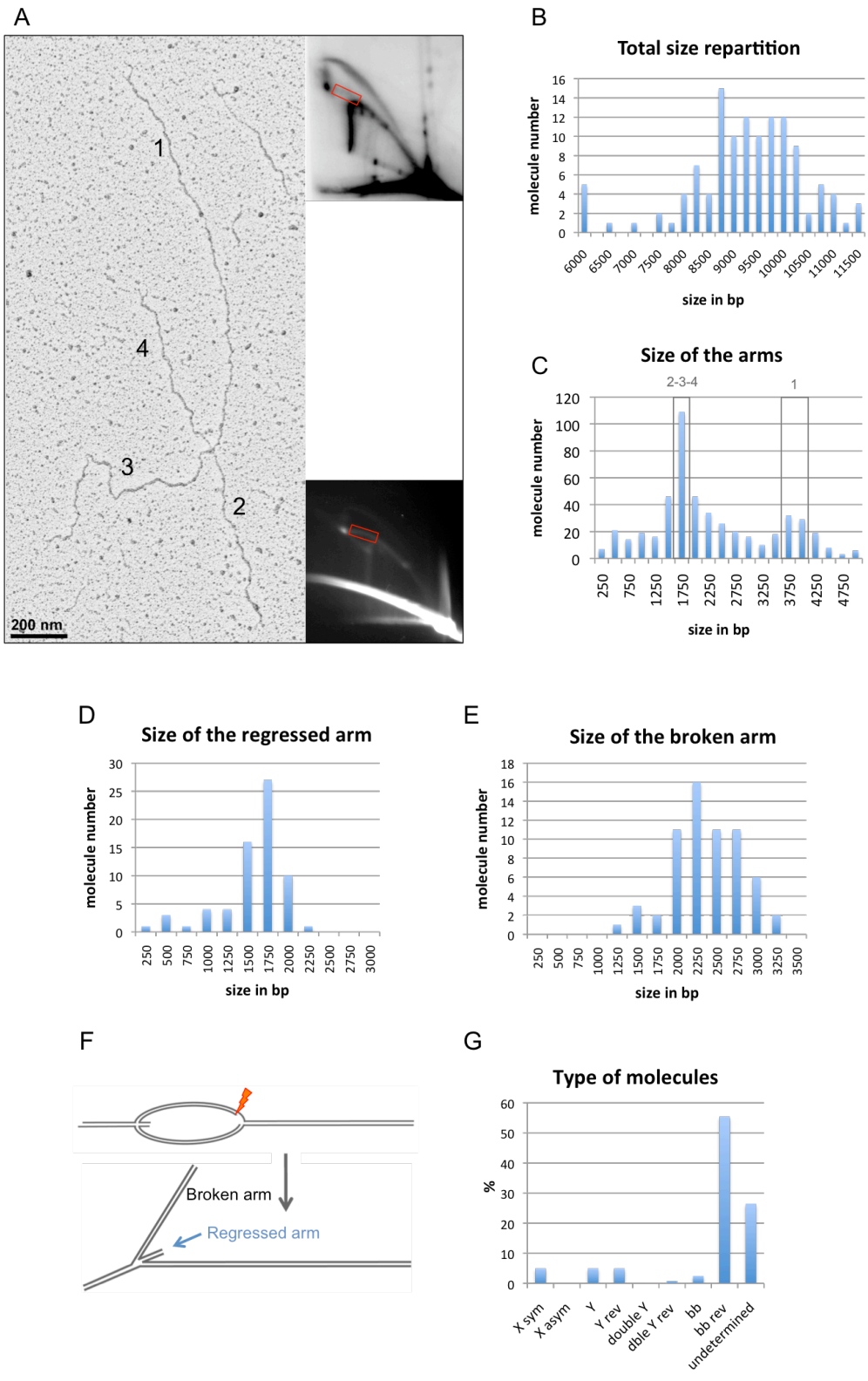


Figure 27. EM analysis of the molecules migrating between the inflection point of the Y-arc and the X-spot in TTC90 plasmids. A. Representative electron micrograph of a molecule migrating in this zone (magnification 46kx). Next to it, typical 2D-gel pattern of TTC90 plasmids by Southern blot and EtBr staining of a preparative 2D-gel. The gel portion used for EM analysis is marked by the red label. B. Size repartition of the molecules.C. Each individual arm of the molecules was measured and plotted. The arms corresponding to the different sizes are indicated in grey. D. Size of the regressed arm. E. Size of the broken arm. F. Model of 4-way junction formation due to "reversed broken bubble". G.Type of the molecules migrating in this zone.

2.7. Characterisation of the intermediates under the Y arc

2.7.1. GAA90 intermediates under the Y

As for the former sample, YS is uniformly represented by 4-way molecules with fixed length of each arm, but no internal symmetry (Fig28A). The length of two of the arms (arms 1 and 3; ~3,9kb and ~1.7kb respectively) also reflects the positioning of the junction at expanded GAA repeats (Figs28 B and C). However, in this case, the other two arms also have a rather homogenous length, of 2.7kb (arm 2) and ~300bp (arm 4) respectively (Figs28 C, D, E). As for the data in 2.6.1, we favour the interpretation of fork breakage at a bubble with the other fork reversed, with arm 2 representing the broken arm of the bubble and arm 4 representing the regressed arm (Fig28F). The data shown in Fig16 also support the broken bubble model. The fact that the YS observed with the EcoRI digestion is visualized as an early Y intermediate with the EcoNI-SacII digestion indeed suggests that the unusual migration in the YS has to be attributed to breakage of the right moving fork, i.e. the one that does not cross the repeats (see model in Fig28G).

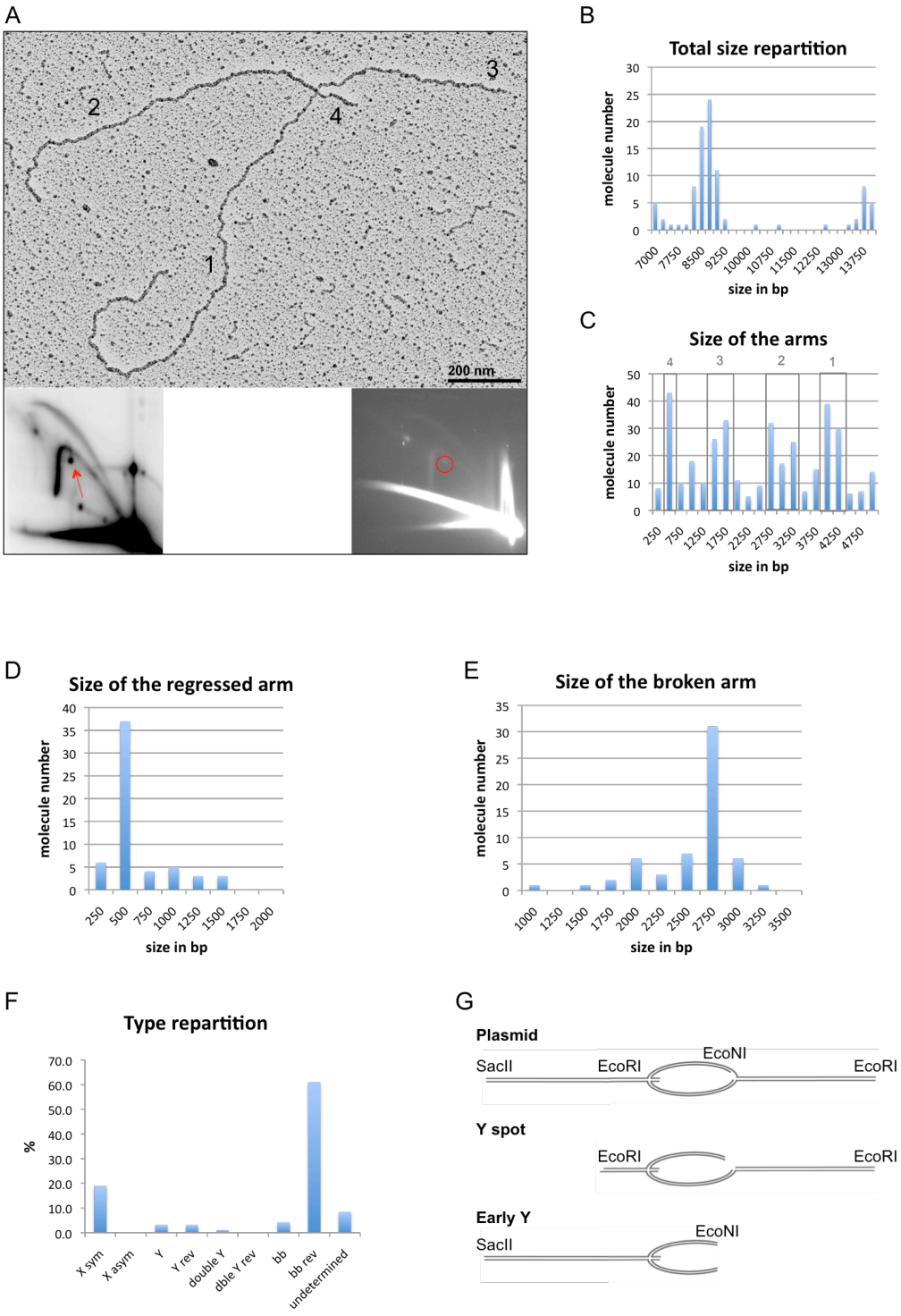


Figure 28. EM analysis of the molecules migrating under the Y-arc in GAA90 plasmids. A. Representative electron micrograph of a molecule migrating in this zone (magnification 46kx). Next to it, typical 2D-gel pattern of GAA90 plasmids by Southern blot and EtBr staining of a preparative 2D-gel. The gel portion used for EM analysis is marked by the red label. B. Size repartition of the molecules. C. Each individual arm of the molecules were measured and blotted. The arms corresponding to the different sizes are indicated in grey. D. Size of the regressed arm. E. Size of the broken arm. F. Type of the molecules migrating in this zone. G. Model explaining the conversion of a broken bubble into a Y according to the digestion.

2.7.2. TTC90 intermediates under the Y-arc (SYS)

Similarly to 2.7.1, 80% of the intermediates migrating under the TTC90 Y arc are consistent with the interpretation of broken bubbles (Figs29 A and D), out of which 70% are reversed. The repartition of arm size (Fig29C) indicates once more that the fork pauses and reverses at the repeats. In this sub-population, the regressed arm measured up to 750bp (Fig29E), while the broken arm of the bubble measured 750-1500bp (Fig29F).

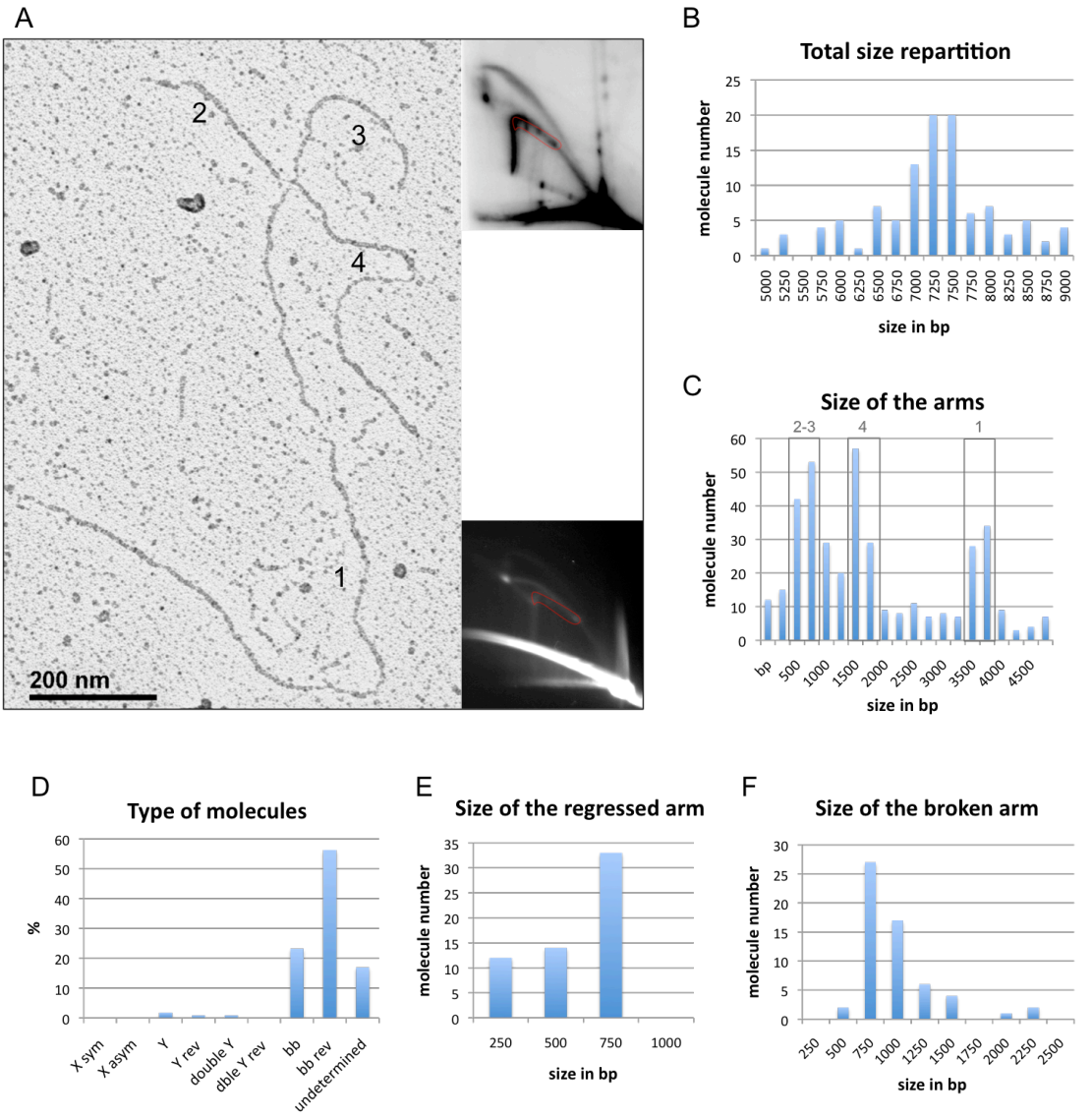


Figure 29. EM analysis of the molecules migrating under the Y-arc by 2D-gel in TTC90 plasmids. A. Representative electron micrograph of a molecule migrating in this zone (magnification 46kx). Next to it, typical 2D-gel pattern of TTC90 plasmids by Southern blot and EtBr staining of a preparative 2D-gel. The gel portion used for EM analysis is marked by the red label. B. Size repartition of the molecules. C. Each individual arm of the molecules were measured and blotted. The arms corresponding to the different sizes are indicated in grey. D. Type of the molecules migrating in this zone. E. Size of the regressed arm. F. Size of the broken arm.

2.8. Characterisation of the spot above the monomer spot

Analysis of the spot migrating above the monomer spot in GAA90 revealed linear molecules of the size of the linearized plasmid (5.6kb) (Figs30 A and B). About 30% of these molecules have detectable single strand gaps (Fig30A), the majority of which is located at a position compatible with the GAA repeats (Fig30C). DNA flaps, defined as ssDNA or dsDNA fragments departing from the linearized plasmid, are also visible. Their position along the plasmid is more random, with a small peak around 1kb (Fig30D). The presence of single strand gaps or flaps could explain the slow migration of the 1n molecule during the second dimension of the 2D-gel.

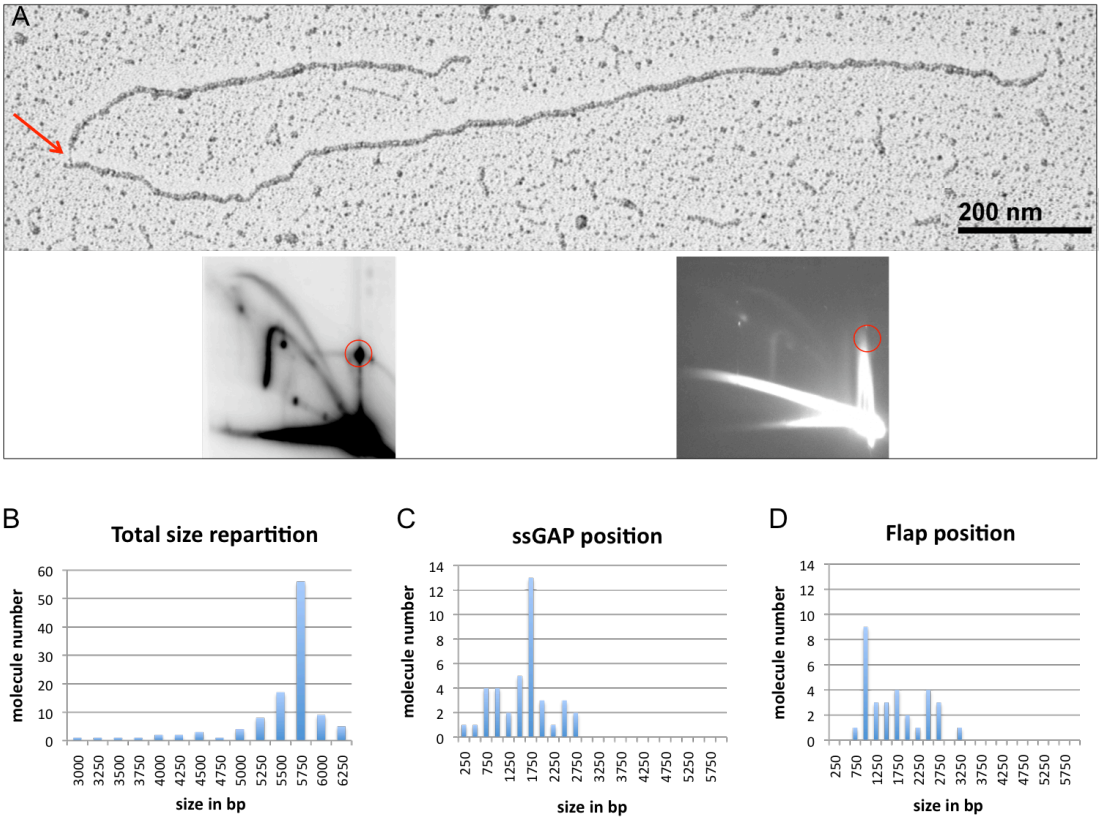
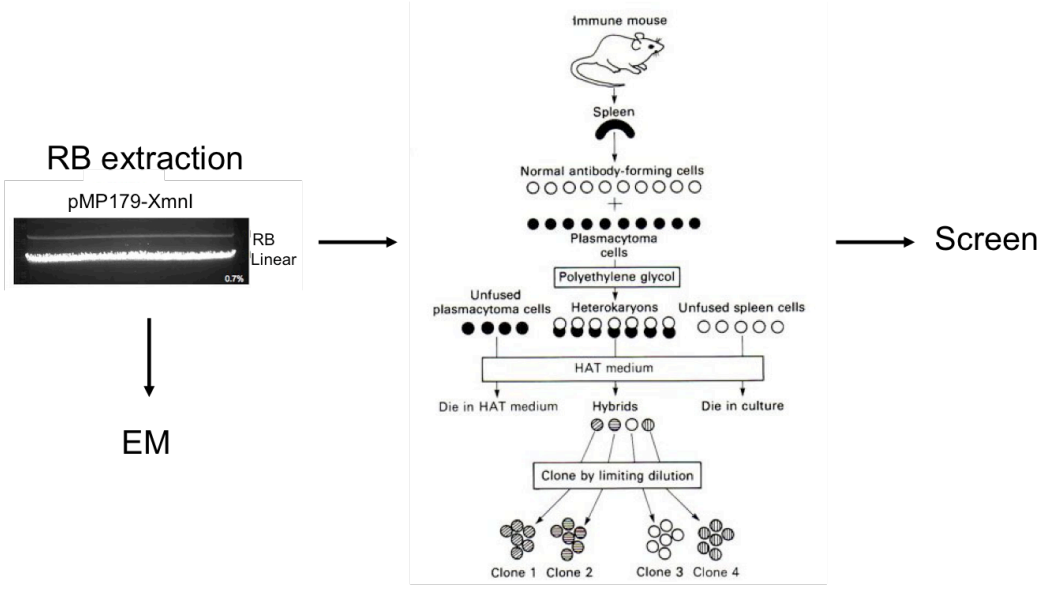


Figure 30. EM analysis of the molecules migrating above the monomer spot in GAA90 plasmids. A. Representative electron micrograph of a molecule migrating in this zone (magnification 46kx). Next to it, typical 2D-gel pattern of GAA90 plasmids by Southern blot and EtBr staining of a preparative 2D-gel. The gel portion used for EM analysis is marked by the red label.B. Size repartition of the molecules. C. Position of single strand gaps measured from the closest extremity of the molecule. D. Position of flaps measured from the closest extremity of the molecule.

3. Results II

The study of TNR is limited by the absence of a simple read-out to follow repeat expansion and contraction. Nowadays the best techniques to determine the size fluctuation are the PCR, which is quite limited in the measurement of long repetitive tracts due to polymerase slippage, and the Southern blot, which is more tedious and time-consuming. Due to these limitations, the study of genetic factors possibly involved in TNR genomic instability has been typically restricted to "educated guess" and single candidate approaches.

To overcome this limitation, we set out to develop a new method to assess - with a simple read-out - not only the size of the repeats, but also their tendency towards instability, by adopting non-B DNA structures (see Introduction). To reach this goal, we aimed to develop an antibody specific for DNA secondary structures arising specifically at expanded GAA repeats. As one of these structures (see below) can be identified by its retarded migration in an agarose gel, we set out to purify the structured DNA from the gel and to inject it in mice as an antigen, to produce specific antibodies. The long term goal of this project would be to use this tool as a read-out in immunofluorescent studies, in order to screen genome-wide for human factors limiting GAA expansion and genomic instability in general at repetitive sequences (Fig31).



Collaboration with Prof. Toshio Mori

Figure 31. Work-flow for the production of an antibody specific for secondary DNA structures at expanded GAA repeats.

3.1. Isolation of a GAA specific secondary structure

The first step in the production of such an antibody is to get an antigen, meaning a population of DNA molecules enriched for alternative secondary structures at expanded GAA repeats. Based on the data shown in Sakamoto et al. 1999, bacterial plasmids containing a number of GAA repeats larger than 50 form a secondary structure, which runs slower than linear DNA in an agarose gel. This retarded band (RB) was identified as two plasmids joint at the level of the repeats in a length- and orientation-dependent manner. Further analysis of this specific RB by our EM protocol provided unprecedented insight in the molecular architecture of these junctions, showing in many cases that the two plasmids are connected at the level of the repeats by the interaction of a single strand on one molecule with the double helix of the second one (Fig32). This evidence supports the previous interpretation of "triplex DNA", that had been mostly based on biochemical assays (Sakamoto et al., 1999).

These data show that gel electrophoresis of GAA containing plasmids amplified *in vivo* in bacteria allows in principle the isolation of secondary structure specific for expanded GAA repeats. We thus proceeded to large scale isolation of this RB using a plasmid containing GAA115 repeats, kindly provided by Massimo Pandolfo. As mice immunization typically requires up to 200ug of antigen and the detected fraction of retarded band is only 3-5% of the total population of plasmid molecules, we had to perform multiple, maxi-preparations of plasmid DNA to reach the 10mg starting material required for the project. After gel migration the RB was extracted, eletro-eluted and concentrated. This procedure, repeated a number of times, due to the limited amount of plasmid DNA that can be loaded in a single preparative gel, finally allowed to obtain the required amount of DNA for immunization, enriching the secondary structure from 3-5% up to 60% (Fig33). We suspect that the residual linear DNA observed after electroelution of the RB is due to *in vitro* resolution of the joint molecules during the procedure.

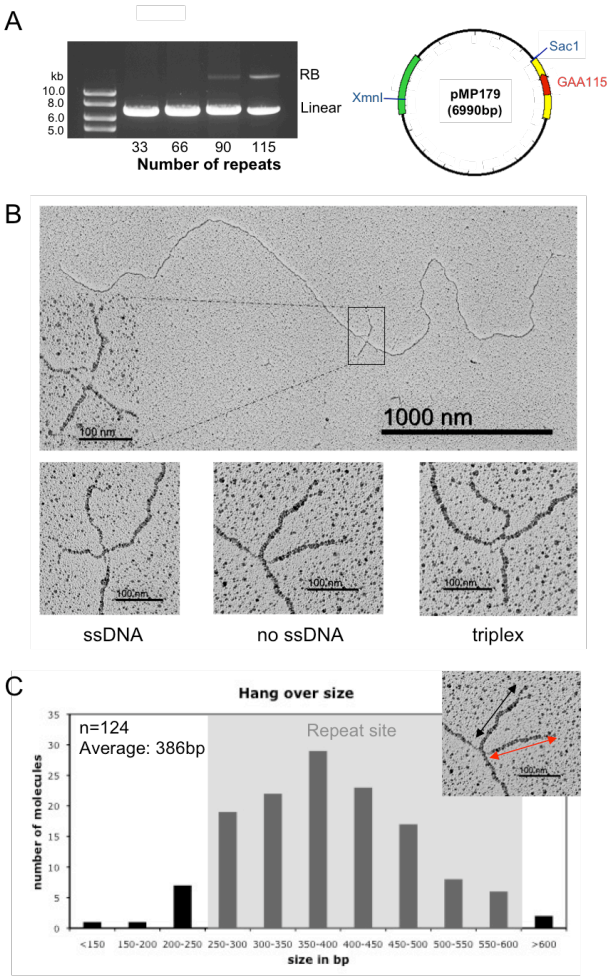


Figure 32. Plasmids containing >50 GAA repeats form a secondary structure in *E.coli*. A. Agarose gel of plasmids containing GAA_n repeats, linearised by SacI, and GAA_{115}-plasmid map (RB= retarded band). B. EM visualisation of GAA_{115}-RB (Magnification : 46kx for the larger picture and 180kx for the smaller insert). C. The position of the junction along the plasmid has been determined after linearization with SacI and contour length measurement of the short arms emanating from it. We found that it coincides exactly with the plasmid location of the repetitive GAA sequences.

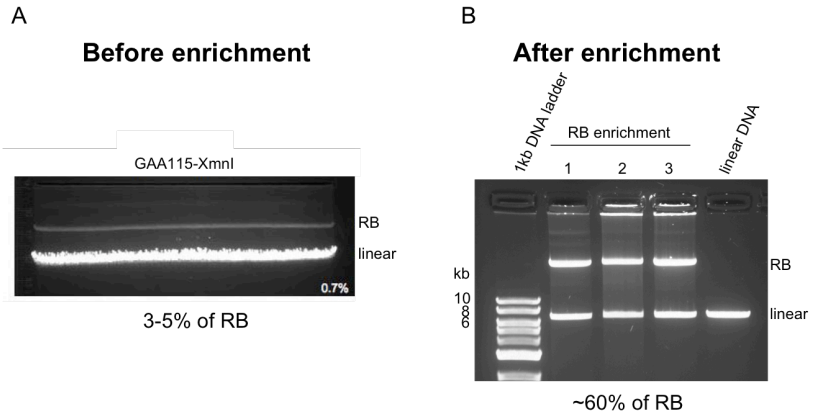


Figure 33. RB enrichment. A. Agarose gel of GAA115 plasmids digested by XmnI. The upper band corresponds to RB and the lower one to the linear DNA (RB/total DNA=3-5%). B. After extraction, electro-elution and concentration, RB was reloaded on an agarose gel to determine the level of enrichment (RB/total DNA=~60%).

3.2. Production of antibodies specific for the secondary structure at expanded GAA repeats.

The antigen was then sent to the group of Prof. Toshio Mori in Japan, which is specialized in the production of DNA specific antibody. All the steps from the immunization to the isolation of monoclonal antibodies were performed in this lab. They first injected the mice with our antigen and waited for the induction of antibodies. Spleen cells were then fused to immortalized cells forming hybridomas and cultured until the formation of colonies. The supernatant of each individual colony was then tested for its specificity in antigen binding by ELISA. Only 3 antibodies (α-RB) out of 280 clones were identified as positive (C23, N49, T57). They show a high specificity for the GAA secondary structure (antigen, RB), a reduced affinity for plasmids containing GAA repeats without structure (linear), and low background binding to random DNA sequence (calf thymus DNA). Comparing the binding intensity (by a colorimetric read-out) of the three candidates to the antigen, T57 shows the highest affinity (Fig34A). Moreover, the low signal detected with the linear (non-structured) plasmid molecules (Figs34 A and B) proves that our antibodies can recognize the DNA structure already present, but cannot induce their formation on non-structured repetitive DNA. The supernatants of these antibodies were then concentrated and sent back to Zurich for further characterization. However, it is important to mention that the final concentration of α-RB antibody is difficult to determine, as the supernatants also contain a large excess of BSA and other contaminant proteins.

In conclusion, the immunization of mice with GAA secondary structure gave rise to 3 specific antibodies in the Mori lab, which could be further analysed by our work in Zurich.

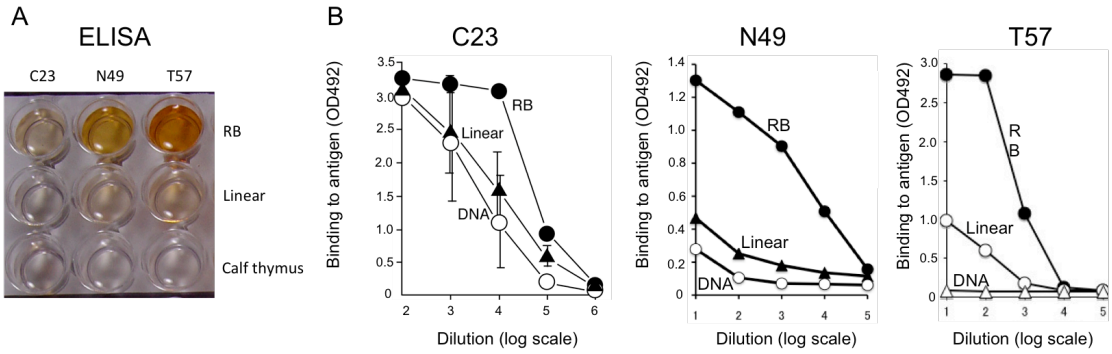


Figure 34. Identification of 3 positive antibodies specific for secondary structures at expanded GAA repeats. A. ELISA assay: the affinity of the antibody is tested on the antigen (RB), GAA115 linear and calf thymus DNA. B. Quantification of ELISA data with different dilutions of the antibodies. (Data from Toshio Mori's lab).

3.3. *In vitro* characterization of α-RB antibodies

3.3.1. Dotblot

The first assay we used to test these DNA antibodies is dot blotting. DNA was spotted on a Zeta probe membrane and then processed as a Western blot: blocking, 1st antibody (α-RB), 2nd antibody (anti-mouse) and chemiluminescence detection. Once the antibody sits on the DNA, the risk of denaturation is reduced and a detergent can be used for secondary antibody and washing. The experimental conditions were established using a general DNA binding antibody provided by Prof. Toshio Mori (data not shown). To test their specificity, the RB-DNA and the same DNA without secondary structure (linear) were blotted and processed for immunodetection using the different α-RB antibodies. This control was made on two different plasmids containing GAA115. The first plasmid is the one used for the antigen production; the second is based on the SV40-pGEM plasmid described in ResultsI (Fig14). For both constructs the results were similar to those already obtained by ELISA (Fig34): T57 is more specific and has higher affinity than P33 and N49 in binding to the secondary structure at expanded GAA repeats (Fig35A). In light of these results, T57 was selected for all following studies. In respect to ELISA, this assay offers several important advantages: it is faster, less laborious and more suitable to test a high number of conditions at the same time.

We thus used this approach to investigate the affinity of the T57 antibody in more detail, starting from a titration curve. The blot shows that, under our

experimental conditions, the antibody can detect less than 10ng of DNA (Fig35B). We then compared its binding to the purified antigen (50ng) and the binding to the total population (1ug) from which it was isolated (Fig35C). The amount of DNA for the two samples was calculated in a way that both contain the same quantity of secondary structure (50ng = 5% of 1ug = RB/total DNA). While a signal is clearly visible for the pure RB sample, no signal can be detected for the total population. This surprising result led to the hypothesis that the RB-specific antibody looses its affinity when the antigen is embedded in an excess of non-antigenic DNA. To address this point more directly, we loaded 50ng of RB mixed with different amounts of competing random DNA (GAA0-XMN1) (Fig35D). The results show an inverted correlation between the binding of the antibody and the quantity of competing DNA (GAA0), substantiating the conclusion that the α-RB antibody has much reduced efficiency when the secondary structure is mixed with other DNA.

This set of data confirmed the specificity of T57 for the GAA secondary structure, but also shows that its affinity is highly reduced when the antigen is not pure.

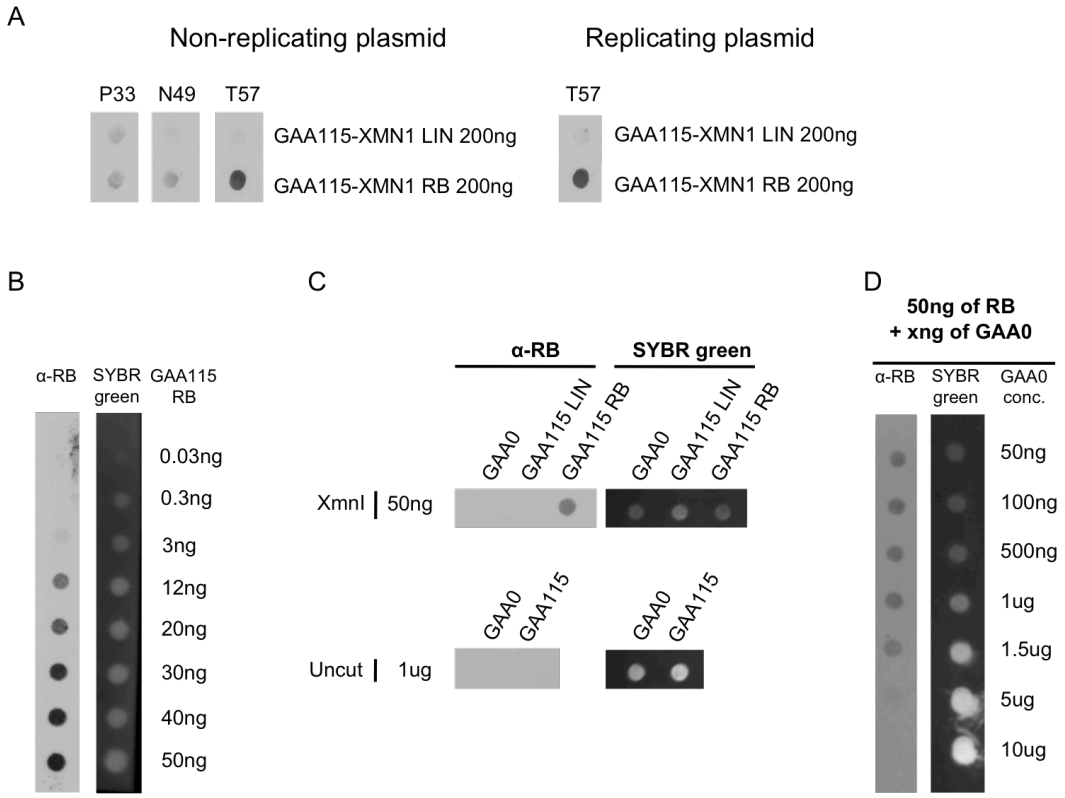


Figure 35. Dot blot assay for antibody-DNA binding. A. α-RB antibody specificity tested by dotblot. T57 antibody was tested on two different types of plasmids, the first one cannot replicate *in vivo,* while the second can. B. Sensitivity of T57 antibody. C. Detection of RB from not enriched GAA115 plasmids (uncut 1ug) by T57. D. Sensitivity of T57 when the antigen is embedded in different concentrations of GAA0 plasmid. (Data from Judith Oehler)

As the antibody can recognize the secondary structure formed at expanded GAA repeats, I aimed to test if the epitope identified was specific. To this purpose we tested oligonucleotides forming different kind of DNA structures: splayed arm, 3'flap, 5'flap, fork, hairpin, D-loop 5', D-loop 3', holliday junction. None of them were detected by T57 (Fig36). However, as the different DNA structure preps still contain a significant fraction of unfolded DNA, we cannot formally exclude that, similar to what shown above for RB itself, the absence of signal is due to the inability of the α-RB antibodies to detect a potential antigen in presence of competing DNA,

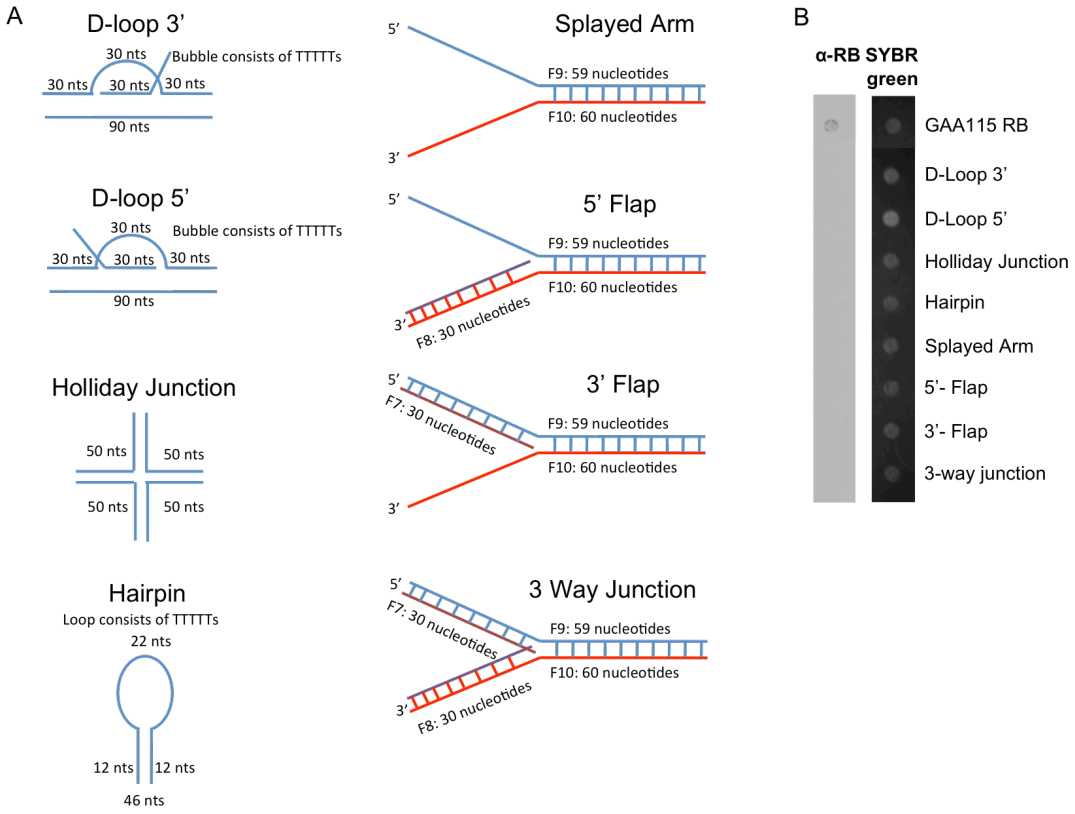


Figure 36. T57 antibody specificity for different secondary structures. A. Description of the tested structures. B. Specificity of T57 for the different structures (180ng) tested by dotblot. (Data from Judith Oehler)

3.3.2. Detection of the secondary stucture by the mean of α-RB after Southern blot

The ability of T57 to detect the GAA secondary structure once transferred to a membrane opened new possibilities, such as attempting to detect - by western blotting of the 2D gels - which of the specific RIs (described in Results I) could carry the antigenic epitope, thus being recognized by the GAA-specific antibody. Furthermore, we reckoned that such experiment could help clarifying if the GAA structures formed in mammalian cells are the same as the bacterial ones and where they migrate within the 2D-gel pattern.

To this aim, we started by running a standard (one-dimensional) agarose gel with RB enriched DNA. The gel was then UV irradiated in order to break down the DNA before Southern blotting with SSC2x. This alternative blotting procedure was chosen to avoid chemical attacks to the DNA, like incubation in NaOH or HCl, to prevent resolution of the secondary structure and to avoid the formation of Y.R.Y triplexes, which are not physiological. Once the DNA was transferred on the membrane, it was processed for immunodetection as described for dot blotting (Fig37).

Controls were made at different steps. To confirm the efficiency of the transfer we checked the dried gel for residual traces of EtBR signals and stained the Zetaprobe membrane with SYBR gold (data not shown). For the Western part, as for the dotblot, we used a general DNA antibody as 1st antibody, which gave a very good signal, detecting each band of the gel. These controls prove that the conditions used for this experiment allow the detection of DNA transferred from an agarose gel (including the RB) by a general DNA antibody.

We then tested the T57 antibody. We ran a gel with different concentration of RB (160ng, 80ng, 20ng) and the linear DNA (160ng) (Fig37A). Despite the good results shown for dot blotting (Fig35), no signal was detectable after Western blot (Fig37B right). To be sure that the DNA was not accidentally degraded during the second immunodetection, we immunostained the same membrane again with the general DNA antibody and the signal was visible (Fig37B left). One possible explanation for this negative result is that the blotting procedure leads to the resolution of the epitope in the RB structure. Considering that all steps

following the blotting are the same as the one used for the dotblot, it means that the structure must be affected before. As we know that the gel migration preserves the majority of the RB structure (Fig33), the only steps differing from dot blotting are the UV irradiation, the prolonged incubation in 2xSSC for transfer and the capillary transfer procedure itself. To identify what leads to RB "denaturation", we tested the individual contributions of these factors by dot blotting (Fig37C). RB and linear DNA were UV irradiated for 6min and incubated with H_2O or 2xSSC overnight. The samples were then blotted on Zetaprob membranes and analysed by immunodetection. Apart from NaOH, none of the reagents used for the transfer affects RB structure, which may suggest a role of the capillary transfer in RB resolution.

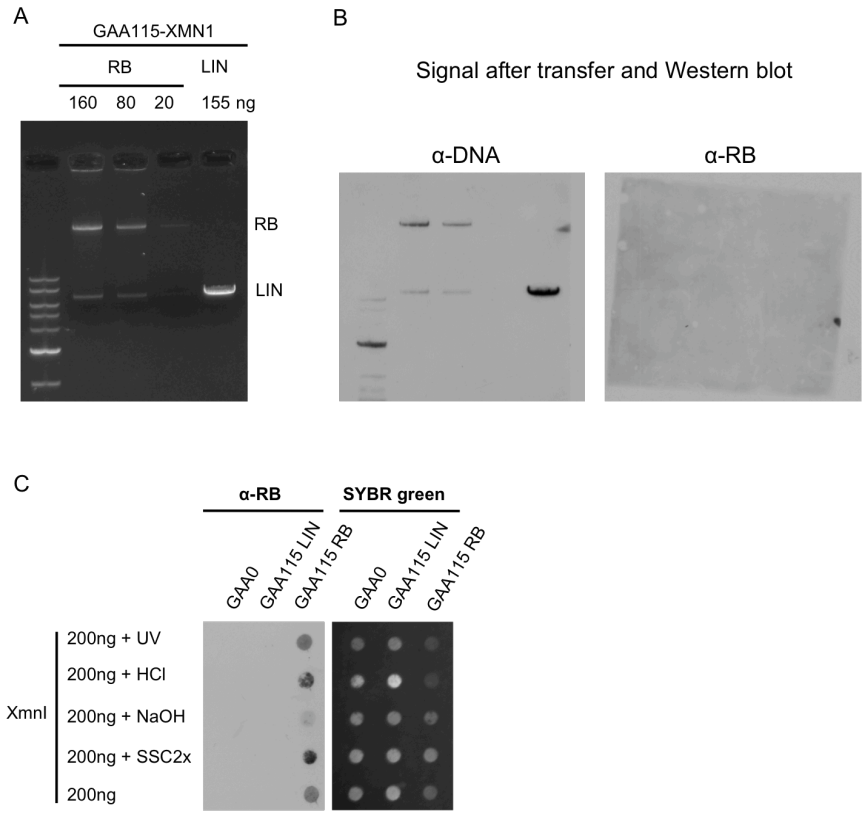


Figure 37. Attempts of RB detection by agarose gel transfer and immunodetection by T57. A. Agarose gel before transfer. B. After Southern blot the membranes were blotted with α-RB or α-DNA antibody. C. Effect of different reagents on RB stability. (Data from Judith Oehler)

3.3.3. *In vitro* transcription assay

GAA repeats are known to reduce the transcription efficiency of the frataxin gene, when they reach a certain number (>50) corresponding to the appearance of the secondary structure (Grabczyk and Usdin, 2000b). This is the known causative event of Friedreich's Ataxia (Pandolfo, 2009). It would thus be of major clinical relevance to find a way to prevent locally non B-DNA formation or to assist the transcription process over it, in order to restore frataxin expression.

Recently published data (Law et al.) showed that a protein named ATR-X can specifically bind repetitive sequences and improve the transcription of genes containing G-quadruplex. These results represent a proof-of-principle that the specific binding of a protein to repetitive DNA may prevent the formation or assist transcription at potential non B-DNA structures, either by causing the recruitment of specific "repair" pathways, by destabilizing the structure or by preventing its formation.

Combining this evidence with the promising data on RB specific recognition by T57, we decided to test whether the GAA secondary structure antibody could assist transcription at expanded GAA tracts. To test this hypothesis we obtained from Prof. Grabczyck different plasmids previously used for *in vitro* transcription assays on GAA repeats (Grabczyk and Usdin, 2000b). We reproduced the key experiment of that report, observing a reduction of transcript levels in a GAA-size dependent manner (Fig38A, lanes 3-6). We next tested the transcription efficiency over GAA repeats in the presence of the α-RB antibody, but observed no detectable differences (Fig38A, lanes 9-12), over an ample range of concentrations (Fig38B). This negative results could be due, as already mentioned above, to the weak affinity of the antibody for the antigen when the epitope is not easily accessible or to the too low purity of the antibody. Alternatively, it could truly reflect the absence of any effect on transcription if the antibody is bound to the secondary structures (see Discussion).

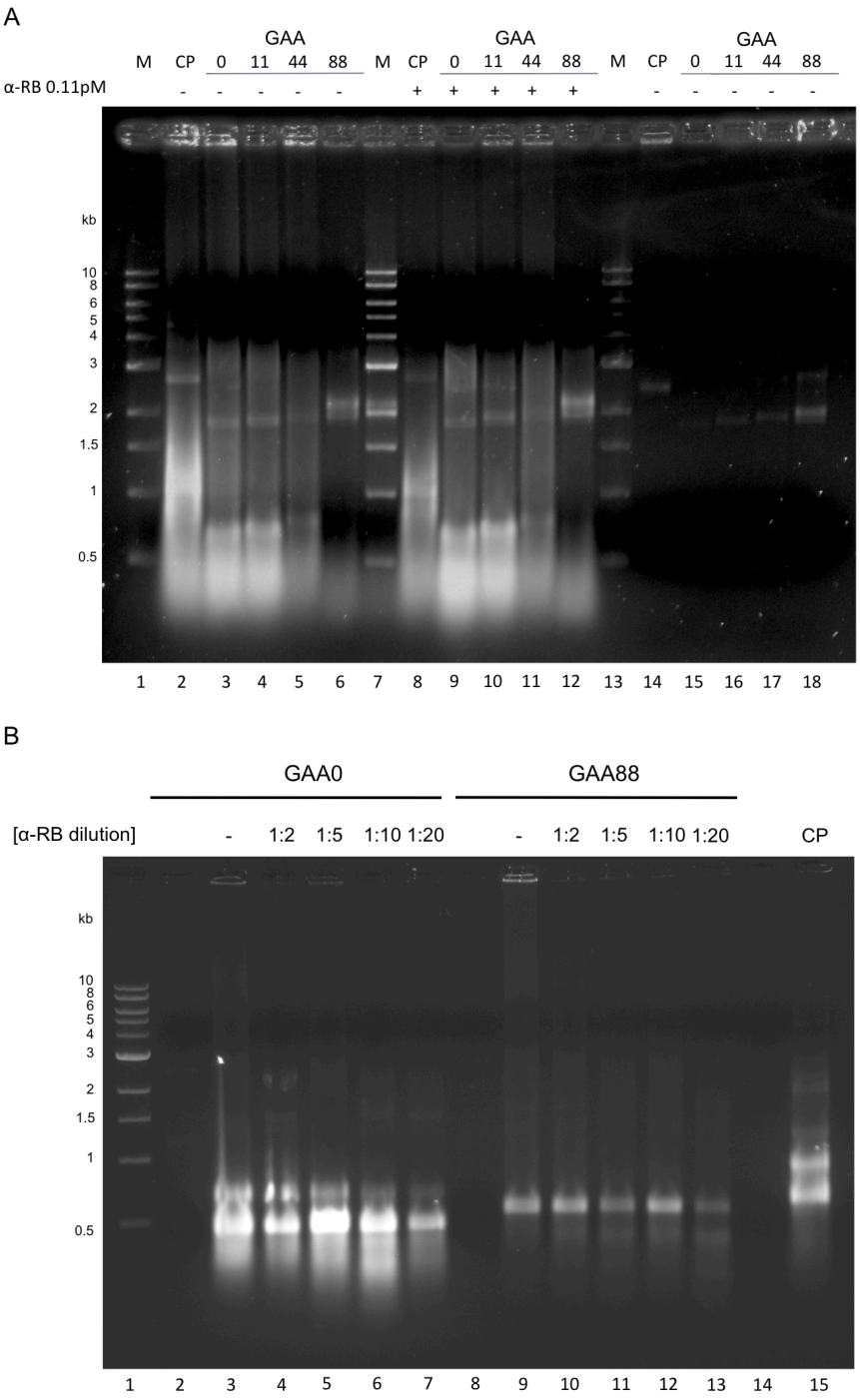


Figure 38. *In vitro* transcription assay on SscI linearised plasmids containing different number of GAA repeats (0,11,44,88). A. *In vitro* transcription assay in absence (lanes 2-6) or in presence (lanes 8-12) of 0.11pmol of α-RB antibody. Lanes 14-18 show the migration of the linear templates before transcription. B. Titration of α-RB antibody during the *in vitro* transcription assay of GAA0 or GAA88 plasmids. (α-RB antibody stock concentration 1:20)(M=1kb marker; CP= control plasmid used as a positive control for the *in vitro* transcription assay, LITMUS 28iMal StuI linearised) (Data from Judith Oehler)

3.4. *In vivo* detection of secondary structures at expanded GAA repeats by T57

In parallel, we tested the antibody *in vivo*. One of the long term goals of this project was the detection of GAA secondary structure *in vivo* by immuno-fluorescence, which would represent an accurate read-out to study genomic instability at the frataxin gene and could be used, in combination with a siRNA library, to screen genome-wide for novel factors involved in GAA repeat instability.

As a starting, positive control, U2OS cells were transfected with the antigen, which contains the secondary structure (purified GAA115-XMNI RB), or with its linear, "non-structured" counterpart (purified GAA115-XMNI linear). These linearized plasmids cannot replicate in mammalian cells, because they miss the nuclear localization signal, the SV40 origin of replication and the large T antigen. They are therefore expected to localize in the cytoplasm. 16h after transfection the cells were processed for immuno-fluorescence using T57 antibody. The pictures (Fig39) show an increased amount of foci in the cells containing the RB structure compared to the linear control. These results confirm the capacity of our antibody to enter the cell and detect specifically the GAA secondary structure in the cytoplasm.

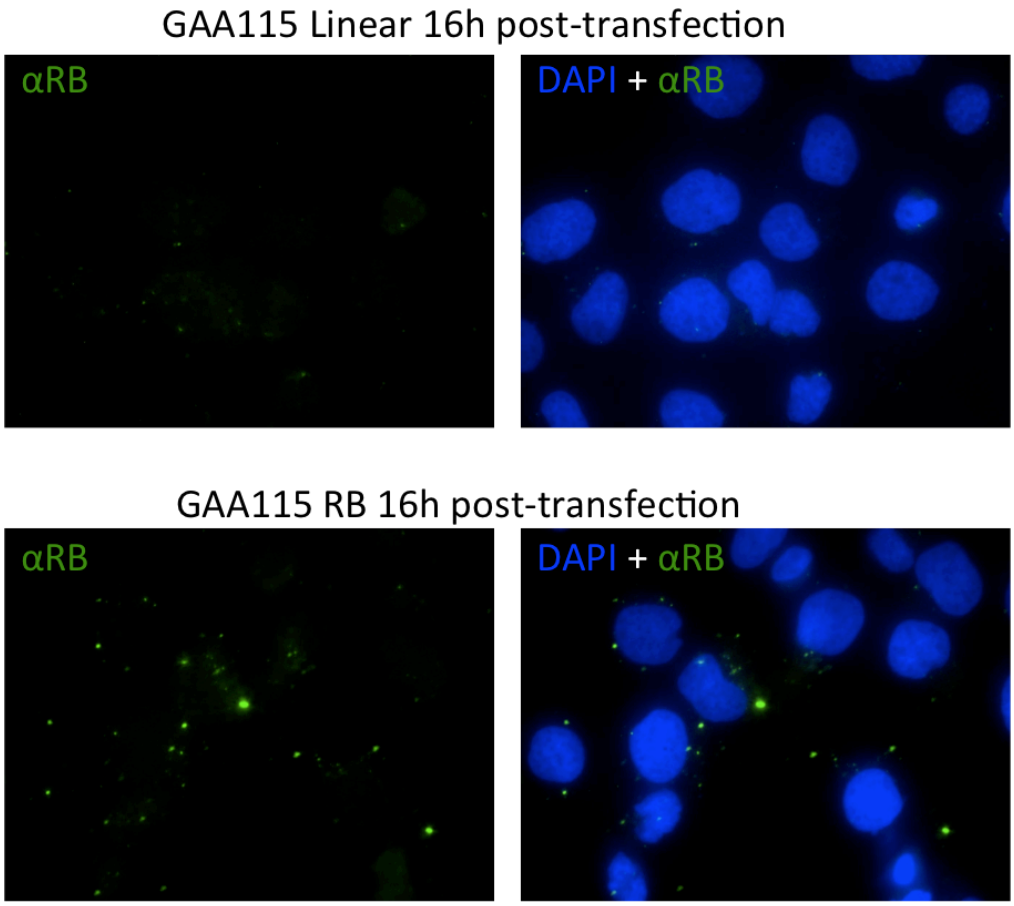


Figure 39. Immuno-fluorescence using α-RB antibody. IF of U2OS cells transfected for 16h with enriched GAA115 linear DNA or GAA115 RB. (αRB=α-RB antibody (green); DAPI (blue)) (Data from Judith Oehler)

Next, we tested T57 on cells transfected with plasmids known to replicate in U2OS and to form GAA-specific secondary structures *in vivo* (see Results I). In order to identify transfected cells we took advantage of the large T antigen expression, which is specifically expressed by our plasmids and can easily be detected by IF. To allow the plasmids to replicate and form the secondary structure the cells were analyzed 48h after transfection. Despite numerous attempts and a large panel of conditions (different fixations, antibody incubations, …) we were not able to detect T57 specific signal in the nucleus (Fig40): the visible signals in Fig40 lower panel do not correspond to transfected cells.

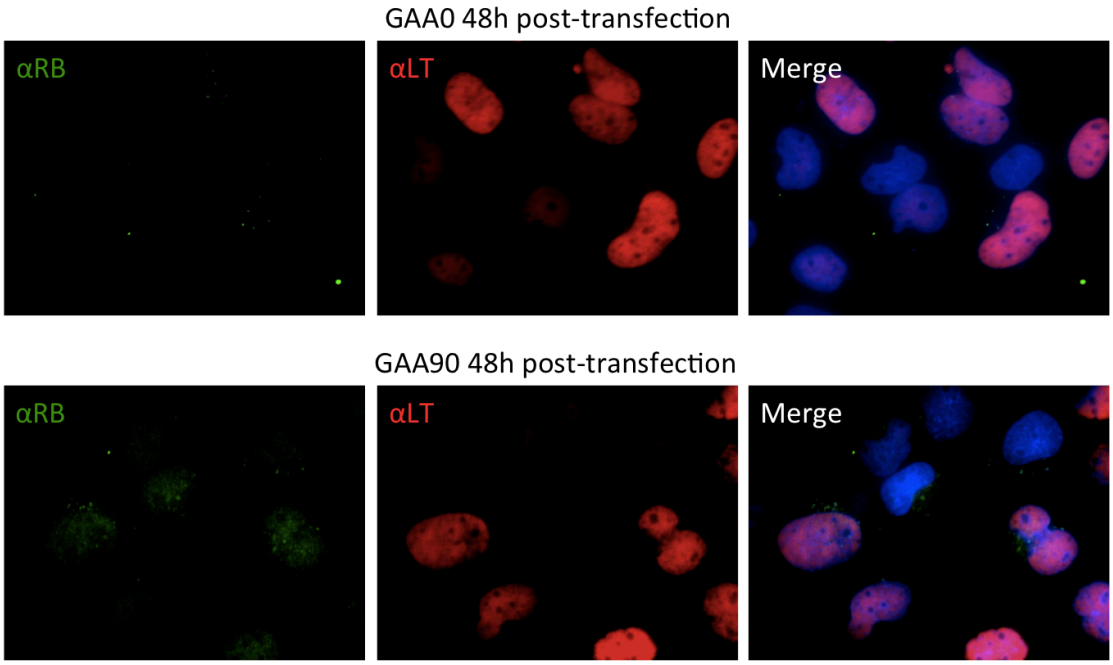


Figure 40. Immuno-fluorescence using α-RB antibody. IF of U2OS cells transfected for 48h with GAA0 or GAA90 plasmids, which can replicate *in vivo*. (αRB=α-RB antibody (green); αLT= large T-antigen antibody (red); DAPI (blue)) (Data from Judith Oehler)

To confirm that our plasmids were indeed directed to the nucleus we also visualized them by fluorescent *in situ* hybridization (FISH) using a probe specific for the plasmids (Fig41). Our results indicate that plasmid-dependent foci signals can be detected both in the nucleus and in the cytoplasm. From these data we can conclude that plasmid DNA enters the nucleus and replicate - in agreement with the data shown in Results I - and should thus be available for detection by the T57 antibody.

This negative result is possibly due to different problems: 1) T57 cannot enter the nucleus; 2) the secondary structure formed *in vivo* in human cells is different from the one formed *in bacteria*; 3) the amount of secondary structure formed *in vivo* is too low to be detected; 4) or most likely, T57 cannot recognize the secondary structure when it is embedded in large excess of non-structure DNA (as already shown by dot blot experiments).

In conclusion, T57 antibody can detect a pre-formed GAA secondary structure transfected in the cytoplasm of U2OS cells, but cannot recognize the one formed *in vivo* in the nucleus, in the context of large excess of genomic DNA.

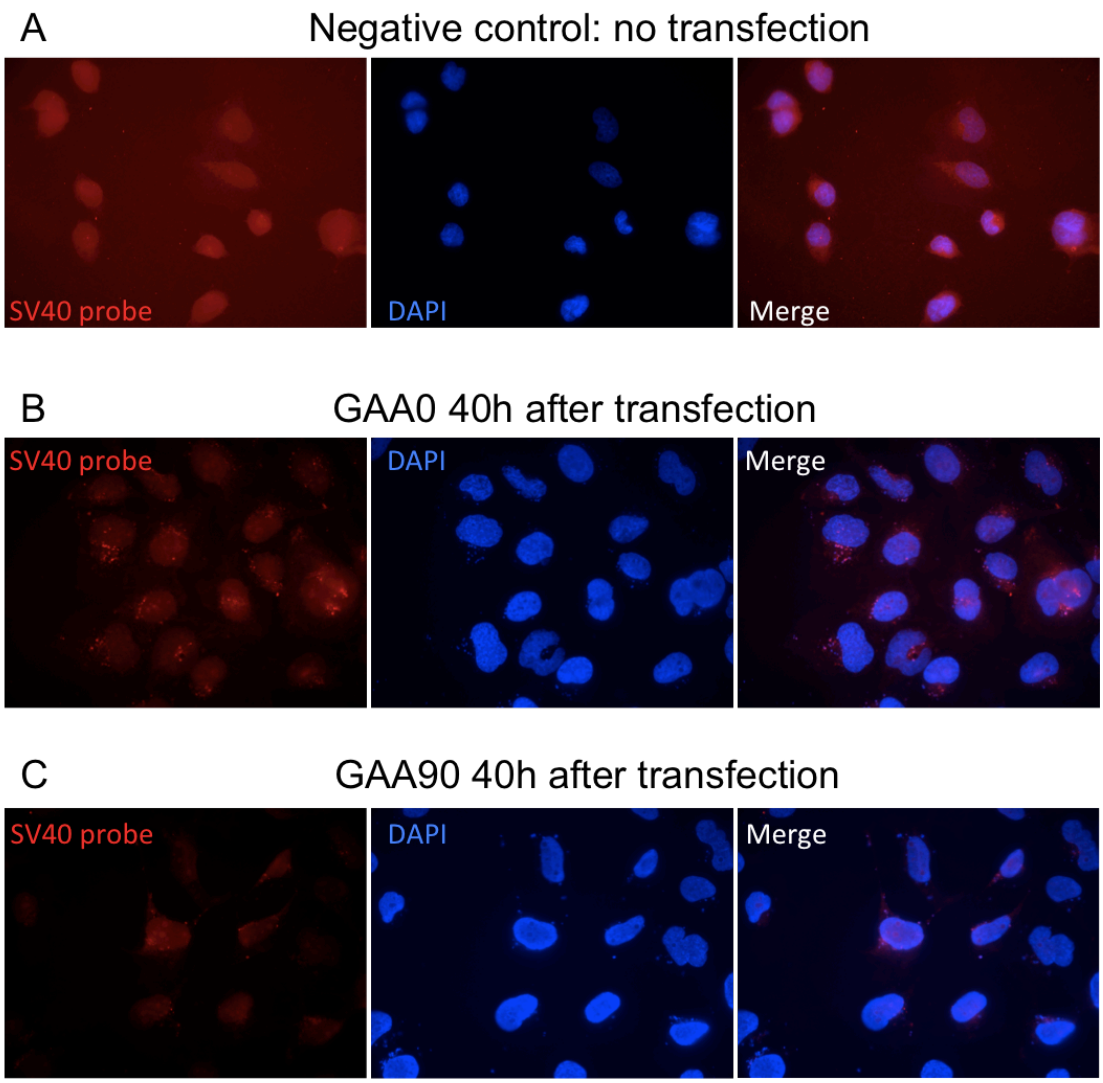


Figure 41. Detection of replicating GAA0 and GAA90 plasmids in U2OS by FISH. A. Not transfected cells (negative control). B. Cells transfected for 40h with GAA0 replicating plasmids. C. Cells transfected for 40h with GAA90 replicating plasmids. (red=SV40 probe; blue=DAPI)

4. Results III: collaborative work

4.1. Exo1 Competes with Repair Synthesis, Converts NER Intermediates to Long ssDNA Gaps, and Promotes Checkpoint Activation (*Molecular Cell* (2010) 40:1-13)

Michele Giannattasio[1*], Cindy Follonier[3,4], Hélène Tourrière[2,4], Fabio Puddu[1], Federico Lazzaro[1], Philippe Pasero[2], Massimo Lopes[3], Paolo Plevani[1*] and Marco Muzi-Falconi[1*]

1. Dipartimento di Scienze Biomolecolari e Biotecnologie, Universita' degli Studi di Milano, Milano, 20133 Italy.

2. IGH - Institute of Human Genetics, CNRS UPR 1142, F-34396 Montpellier, France

3. Institute of Molecular Cancer Research, University of Zuerich, Zuerich, CH-8057, Switzerland

4. These authors contributed equally to this work.

Abstract

Ultraviolet (UV) light induces DNA-damage checkpoints and mutagenesis, which are involved in cancer protection and tumorigenesis, respectively. How cells identify DNA lesions and convert them to checkpoint-activating structures is a major question. We show that during repair of UV lesions in noncycling cells, Exo1-mediated processing of nucleotide excision repair (NER) intermediates competes with repair DNA synthesis. Impediments of the refilling reaction allow Exo1 to generate extended ssDNA gaps, detectable by electron microscopy, which drive Mec1 kinase activation and will be refilled by long-patch repair synthesis, as shown by DNA combing. We provide evidence that this mechanism may be stimulated by closely opposing UV lesions, represents a strategy to redirect problematic repair intermediates to alternative repair pathways, and may also be extended to physically different DNA damages. Our work has significant implications for understanding the coordination between repair of DNA lesions and checkpoint pathways to preserve genome stability.

My contribution: Fig4.

I contributed to this paper making extensive EM analysis on yeast genomic DNA. Control or Δexo1 cells were synchronized in G1 and UV irradiated. After genomic extraction and enrichment, the samples were analysed by EM. The data showed direct evidence of increased single strand gaps after UV irradiation and also confirmed the involvement of Exo1 in their processing. These results were of crucial importance to gain insight into this biological mechanism and were thus recognized by a second authorship.

Please cite this article in press as: Giannattasio et al., Exo1 Competes with Repair Synthesis, Converts NER Intermediates to Long ssDNA Gaps, and Promotes Checkpoint Activation, Molecular Cell (2010), doi:10.1016/j.molcel.2010.09.004

Molecular Cell
Article

Exo1 Competes with Repair Synthesis, Converts NER Intermediates to Long ssDNA Gaps, and Promotes Checkpoint Activation

Michele Giannattasio,[1] Cindy Follonier,[3,4] Hélène Tourrière,[2,4] Fabio Puddu,[1] Federico Lazzaro,[1] Philippe Pasero,[2] Massimo Lopes,[3] Paolo Plevani,[1,*] and Marco Muzi-Falconi[1,*]
[1]Dipartimento di Scienze Biomolecolari e Biotecnologie, Università degli Studi di Milano, Milano, 20133, Italy
[2]Institute of Human Genetics (IGH), CNRS UPR 1142, F-34396 Montpellier, France
[3]Institute of Molecular Cancer Research, University of Zürich, Zürich, CH-8057, Switzerland
[4]These authors contributed equally to this work
*Correspondence: paolo.plevani@unimi.it (P.P.), marco.muzifalconi@unimi.it (M.M.-F.)
DOI 10.1016/j.molcel.2010.09.004

SUMMARY

Ultraviolet (UV) light induces DNA-damage checkpoints and mutagenesis, which are involved in cancer protection and tumorigenesis, respectively. How cells identify DNA lesions and convert them to checkpoint-activating structures is a major question. We show that during repair of UV lesions in noncycling cells, Exo1-mediated processing of nucleotide excision repair (NER) intermediates competes with repair DNA synthesis. Impediments of the refilling reaction allow Exo1 to generate extended ssDNA gaps, detectable by electron microscopy, which drive Mec1 kinase activation and will be refilled by long-patch repair synthesis, as shown by DNA combing. We provide evidence that this mechanism may be stimulated by closely opposing UV lesions, represents a strategy to redirect problematic repair intermediates to alternative repair pathways, and may also be extended to physically different DNA damages. Our work has significant implications for understanding the coordination between repair of DNA lesions and checkpoint pathways to preserve genome stability.

INTRODUCTION

Ultraviolet (UV) radiation damages DNA, causing base modifications, mostly cyclobutane pyrimidine dimers (CPD) and 6,4 photoproducts (64-PP) (Friedberg et al., 2006), that are responsible for UV-induced mutagenesis and are responsible for the pathogenetic effects of sunlight.

In mammalian cells, UV-induced DNA damage can only be repaired by nucleotide excision repair (NER), through incisions of the damaged DNA strand 5′ and 3′ to the adduct, and by removal of an oligonucleotide containing the lesion. The ~30 nt long ssDNA gap is then refilled by DNA polymerases, which copy the nondamaged strand, and sealed by DNA ligase. The clinical relevance of NER is underlined by the existence of syndromes (e.g., xeroderma pigmentosum, Cockayne syndrome) caused by mutations affecting genes coding for NER factors (Friedberg et al., 2006).

If NER cannot effectively fix damaged DNA, lesions persist through the cell cycle and interfere with replicative DNA synthesis, leading to the formation of daughter strand gaps, which are due to repriming events downstream of the lesion (Rupp and Howard-Flanders, 1968; Cordeiro-Stone et al., 1979; Lehmann, 1979; Lopes et al., 2006; Lehmann and Fuchs, 2006). To cope with damaged DNA, cells have evolved different strategies, known as DNA-damage tolerance mechanisms, which entail postreplication repair (PRR). ssDNA gaps generated as a consequence of replication-blocking lesions can be refilled by translesion DNA synthesis (TLS), via specialized DNA polymerases inserting nucleotides opposite noninstructional or misinstructional DNA lesions (Goodman, 2002; Rattray and Strathern, 2003), or they can be directed toward a template switching pathway (Rupp et al., 1971; Zhang and Lawrence, 2005). Intriguingly, recent work demonstrated an unexpected role for mammalian pol kappa in the refilling step of NER (Ogi and Lehmann, 2006; Ogi et al., 2010).

Exo1 is a member of the Rad2 family of structure-specific nucleases, and it possesses a 5′-3′ exonuclease activity and a 5′-flap endonuclease activity in vitro. Several studies have implicated this enzyme in a variety of DNA metabolic processes (Tran et al., 2004): Exo1 has a fundamental role in mismatch repair and it has been recently involved in processing and tolerance of stalled replication forks (Tran et al., 2004; Cotta-Ramusino et al., 2005; Segurado and Diffley, 2008). Consistently, mouse *EXO1* mutants exhibit a high proneness to tumor development (Wei et al., 2003).

Checkpoints are surveillance mechanisms responsible for the maintenance of genomic stability and cell viability after genotoxic damage (Lazzaro et al., 2009). Loss of checkpoint function leads to chromosomal instability and promotes cancerogenesis (Kerzendorfer and O'Driscoll, 2009). The mechanism underlying the checkpoint response is a phosphorylation-based signal transduction cascade conserved in all eukaryotes (Lazzaro et al., 2009). The actual structure accountable for signaling DNA damage has been well characterized in the case

Please cite this article in press as: Giannattasio et al., Exo1 Competes with Repair Synthesis, Converts NER Intermediates to Long ssDNA Gaps, and Promotes Checkpoint Activation, Molecular Cell (2010), doi:10.1016/j.molcel.2010.09.004

of double-strand breaks (DSBs), where broken DNA ends are resected 5′ to 3′ to generate ssDNA filaments (Harrison and Haber, 2006). Replication protein A (RPA)-covered ssDNA is thought to recruit the apical checkpoint kinase (Mec1-Ddc2 in budding yeast, ATR-ATRIP in mammalian cells) and the 9-1-1 complex (Rad17-Mec3-Ddc1 in budding yeast, Rad1-Hus1-Rad9 in mammalian cells), triggering the signal transduction cascade (Zou and Elledge, 2003). As a consequence of this activation, Mec1 phosphorylates, directly or indirectly, a number of factors (e.g., Ddc2, H2A, Ddc1, Rad9, Rad53, etc.), and this event is generally used to follow the signal throughout the cascade.

In cycling eukaryotic cells, bulky lesions induced by UV light will block progression of the replication fork, leading to checkpoint activation (Ward et al., 2004). How UV lesions lead to a similar checkpoint response in the absence of DNA replication (e.g., nonproliferating cells) is less understood and is the object of the present study.

UV lesions are not sufficient to directly trigger a prompt DNA-damage response in noncycling cells. Indeed, both in cell-cycle-arrested yeast cells and in resting human fibroblasts, activation of the checkpoint induced by UV irradiation requires a functional nucleotide excision apparatus (Giannattasio et al., 2004; Marini et al., 2006; Marti et al., 2006). These findings suggest that NER factors may be involved in directly recruiting checkpoint proteins to damage sites and/or that NER intermediates may be the structure recognized by checkpoint sensors (Matsumoto et al., 2007; Lazzaro et al., 2009).

Here we show that in noncycling cells, activation of the checkpoint after UV irradiation requires the activity of Exo1. We provide physical evidence that in G1 cells, Exo1 competes with the refilling polymerase and, capturing NER intermediates stabilized by impediments in repair synthesis, transforms them into long ssDNA regions, which trigger the DNA-damage checkpoint response. This mechanism represents a way to divert unrepairable intermediates to different repair pathways.

RESULTS

Exo1 Is Required to Respond to UV Damage in Noncycling Cells

The mechanism underlying the response to UV light has only been partially investigated: we have shown that NER-mediated processing of UV lesions is a prerequisite for the rapid response to UV light (Giannattasio et al., 2004; Marini et al., 2006). On the other hand, the gaps generated by NER are in the range of 30 nt in length, quite short to load checkpoint complexes, suggesting that NER intermediates are not the activating structures.

We hypothesized that nuclease activities may be involved in checkpoint activation, generating a larger ssDNA gap during NER processing.

To investigate this possibility, we analyzed the checkpoint response to UV light in several yeast nuclease mutants (Figure S1A, available online), and we report our findings on Exo1. Since NER is only involved in checkpoint activation in noncycling cells, we studied G1- and G2-arrested cells. In Figure 1, G1-arrested *exo1Δ* cells were UV irradiated, and activation of the checkpoint response was verified by monitoring phosphorylation of the checkpoint kinase Rad53. Figure 1A shows that, while in wild-type (WT) cells Rad53 very rapidly reaches the hyperphosphorylated state, indicative of a full-blown DNA-damage checkpoint, cells lacking Exo1 are extremely defective in activating Rad53 and lack a G1 checkpoint. Indeed, similar to a checkpoint-defective *rad9Δmec3Δ* control, *exo1Δ* cells UV irradiated in G1 and allowed to proceed in the cell cycle fail to delay the G1/S phase transition, as seen by following bud emergence (Figure 1B).

We asked whether Exo1 activity in the response to UV light was at the top of the signal transduction cascade, where Mec1 is recruited. Ddc2 interacts with Mec1, and it is a direct substrate for Mec1 kinase. Figure 1C shows that the rapid UV-induced phosphorylation of Ddc2 typical of a WT strain is lost in *exo1Δ* cells. A similar result was obtained for phosphorylation of histone H2A, another direct Mec1 target (Figure 1D), suggesting that Exo1 is involved in recruiting or activating the checkpoint sensors.

To verify the requirement for Exo1 enzymatic activity, we analyzed checkpoint activation with two catalytically inactive mutants, *exo1-D173A* and *exo1-E150D* (Tran et al., 2002). Both mutations impair Rad53 phosphorylation after UV irradiation (Figure 1E), indicating a requirement for the nuclease activity of Exo1. Rad53 phosphorylation in *exo1Δ* strains is also significantly compromised if cells are irradiated after a nocodazole arrest, suggesting a function for Exo1 at the G2/M checkpoint as well (Figure S1B). Exo1 does not play any role in NER; in fact, *exo1Δ* cells are not UV sensitive and remove UV lesions from chromosomes (Figure S1C), indicating that excision of damaged DNA occurs normally. On the other hand, Exo1 is known to play an important function in mismatch repair; we thus investigated a possible role for MMR in UV-induced checkpoint activation. Figure S1D shows that *msh2Δ* and *mlh1Δ* cells do not exhibit any checkpoint defect in our assay. Similarly, the MRX complex does not seem to be involved in any way in this rapid response (Figure S1E), which differs from what was reported for a delayed UV response (Nakada et al., 2004).

All these results suggest that NER processing, albeit essential, is not sufficient for activating the checkpoint response, and that Exo1 nucleolytic activity also plays a relevant role in this pathway downstream of NER.

Exo1 Processes UV-Damaged Chromosomes Generating ssDNA Regions

Generation of ssDNA regions through Exo1-dependent processing of damaged chromosomes may provide a possible explanation for our results. We analyzed chromosomal DNA after UV irradiation by pulsed-field gel electrophoresis (PFGE). Figure 2A shows that UV light alters the structure of chromosomal DNA; indeed, after irradiation a noticeable amount of genomic DNA runs as a ''cloud'' close to the wells instead of running as discrete chromosomes. Such intermediates are readily detectable by ethidium bromide staining and even more clearly visualized by Southern blotting (Figure 2B and Figure S2) and may correspond to ssDNA-containing molecules, which may be retarded because of the flexibility imparted by the ssDNA region (Chen et al., 2007). The formation of these UV-dependent structures requires NER and Exo1 activity, as shown by the reduction in

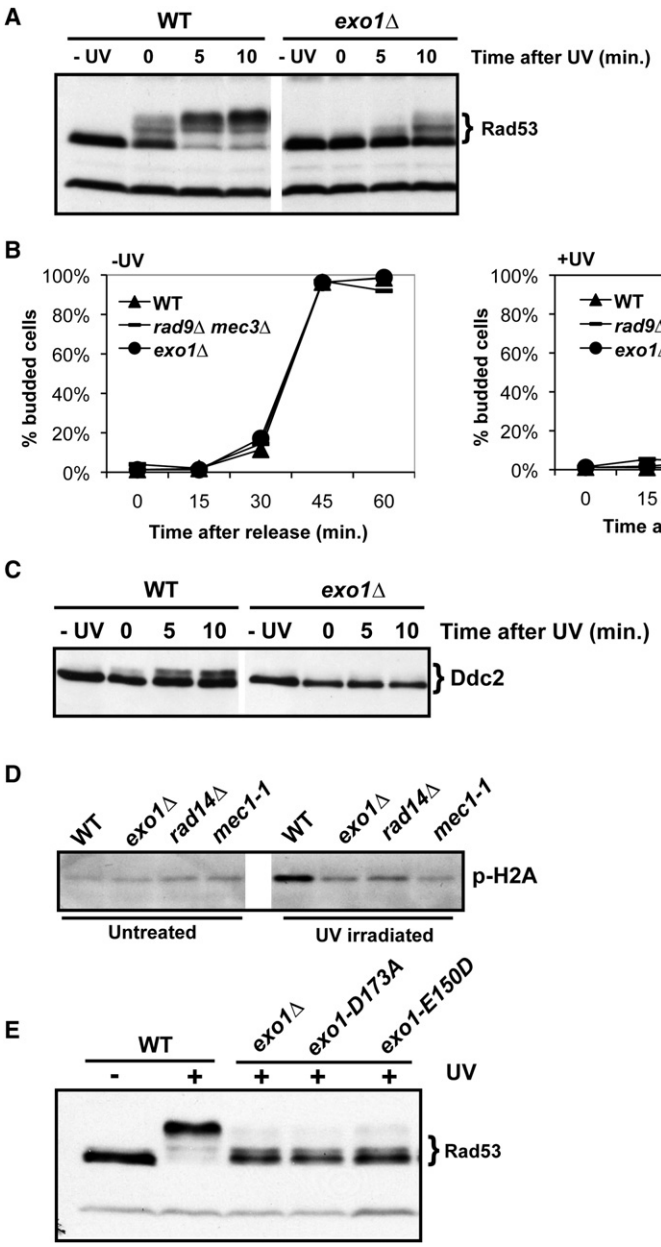


Figure 1. Exonucleolytic Activity of Exo1 Is Necessary for Mec1 Activation and for the G1 Checkpoint after UV Treatment

(A) WT and *exo1*Δ cells were arrested in G1 with α-factor, UV irradiated (75 J/m^2), and held in G1. At the indicated time points, Rad53 phosphorylation was detected by western blotting as a shifted protein species.

(B) The same strains as in (A), plus the *mec3*Δ *rad9*Δ-positive control, were arrested in G1, UV irradiated (40 J/m^2), and released in the cell cycle. Kinetics of bud emergence was measured as described (Giannattasio et al., 2004).

(C) *ddc2HA* and *ddc2HA exo1*Δ strains were treated as in (A) and, at different times after UV irradiation (75 J/m^2), Ddc2 phosphorylation was detected by western blotting.

(D) WT, *exo1*Δ, *rad14*Δ, and *mec1-1 sml1* cells were arrested in G1 and treated as in (A). Histone H2A phosphorylation was detected 30 min after the UV irradiation (75 J/m^2) with a phosphospecific antibody.

(E) *exo1-D173A*, *exo1-E150D*, and the corresponding WT strain were arrested in G1 and treated as in (A). Rad53 activation was detected by western blotting immediately after UV irradiation (75 J/m^2).

the cloud signal in a NER-defective *rad14Δ* strain and in *exo1Δ* cells (Figure 2B).

Two lines of evidence suggest that these unusual intermediates result from UV-induced processing of chromosomes, leading to extended ssDNA gaps. Firstly, BrdU incorporation during repair synthesis of UV-irradiated WT chromosomes is strongly enriched in the cloud signal. On the other hand, in Exo1-defective cells BrdU incorporation is reduced, and it is distributed in all chromosomes, whereas NER-defective cells show no incorporation at all. This observation suggests that Exo1 is processing NER-dependent repair intermediates (Figure 2C). Secondly, incubation of agarose plugs with ssDNA-specific S1 nuclease prior to PFGE virtually eliminated the cloud signal, while cleaved DNA molecules accumulated as a smear at lower molecular weights in the UV-irradiated, S1-treated sample. These ssDNA-containing regions are generated only after UV irradiation, since S1 nuclease has no evident effect on nonirradiated chromosomes (Figure 2D and Figure S2B).

Finally, we analyzed the relationship between the generation of the ssDNA intermediates and Rad53 phosphorylation in a UV dose curve and in a time course of repair. Accumulation of ssDNA structures is proportional to the UV dose and correlates with Rad53 phosphorylation (Figure 2E). Moreover, during repair we observed a time-dependent resolution of ssDNA intermediates, which parallels the reduction in phosphorylated Rad53 (Figure 2F). Altogether, these results suggest that NER- and Exo1-dependent processing of UV-damaged chromosomes generates ssDNA regions that may lead to activation of the apical Mec1 kinase.

Molecular Combing Reveals the Formation of UV- and Exo1-Dependent ssDNA Gaps

To detect in vivo the presence of Exo1-dependent ssDNA gaps, we indirectly measured them by DNA combing after labeling repair synthesis with BrdU. Cultures were arrested in G1, UV irradiated, and immediately supplemented with BrdU. Repair was allowed to proceed for 2 hr in G1. To measure gap refilling, we prepared genomic DNA and processed it for DNA combing.

Please cite this article in press as: Giannattasio et al., Exo1 Competes with Repair Synthesis, Converts NER Intermediates to Long ssDNA Gaps, and Promotes Checkpoint Activation, Molecular Cell (2010), doi:10.1016/j.molcel.2010.09.004

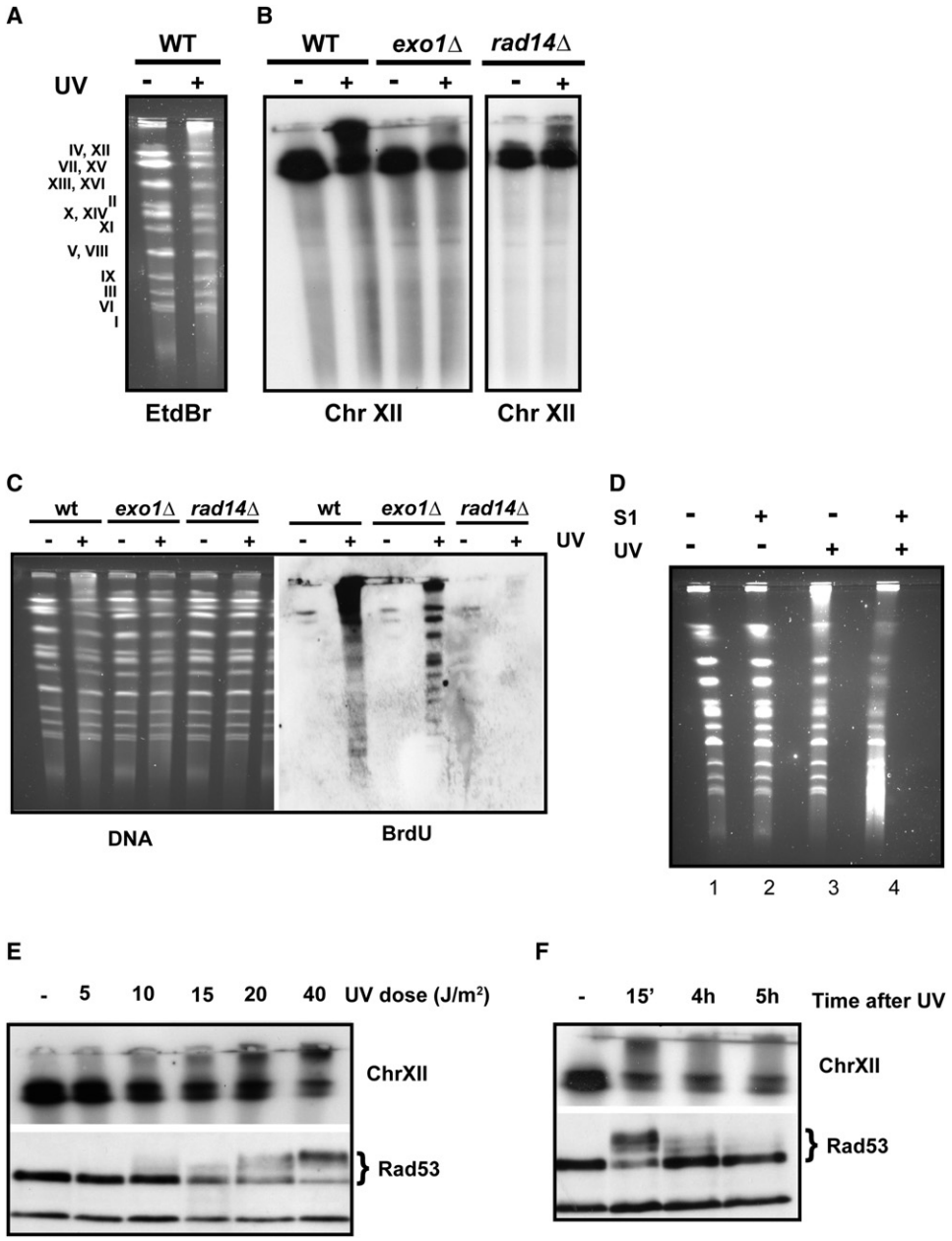


Figure 2. ssDNA Intermediates Generated by NER and Exo1 Correlate with Rad53 Activation

(A) WT, *exo1*Δ, and *rad14*Δ cells were arrested in G1, UV irradiated (75 J/m^2), and held in G1. Chromosomal DNA was prepared 15 min after irradiation and analyzed by PFGE. Chromosomes were visualized by ethidium bromide staining.

(B) Chromosome XII was visualized by Southern blotting.

(C) WT, *exo1*Δ, and *rad14*Δ strains in a *rho*$^\circ$ *bar1*Δ *cdc45*^{td} background were arrested in G1 and shifted to 37°C to inactivate Cdc45. Cells were UV treated (75 J/m^2) and resuspended in prewarmed (37°C) medium containing α-factor and BrdU. After 2 hr of recovery at 37°C, cells were fixed with sodium azide, and UV-induced BrdU incorporation in chromosomal DNA was detected by PFGE and western blotting.

(D) Chromosomal DNA prepared in (A) was treated with S1 nuclease, separated by PFGE, and stained with ethidium bromide.

(E) WT cells were arrested in G1 and treated as in (A) with the indicated UV doses. Rad53 phosphorylation was detected by western blotting, and the accumulation of UV-dependent DNA intermediates was monitored by Southern blotting as in (B).

(F) WT cells were arrested in G1, UV treated (40 J/m^2), and held in G1. At the indicated time points, samples were analyzed for Rad53 phosphorylation and for accumulation of chromosomal intermediates as in (D).

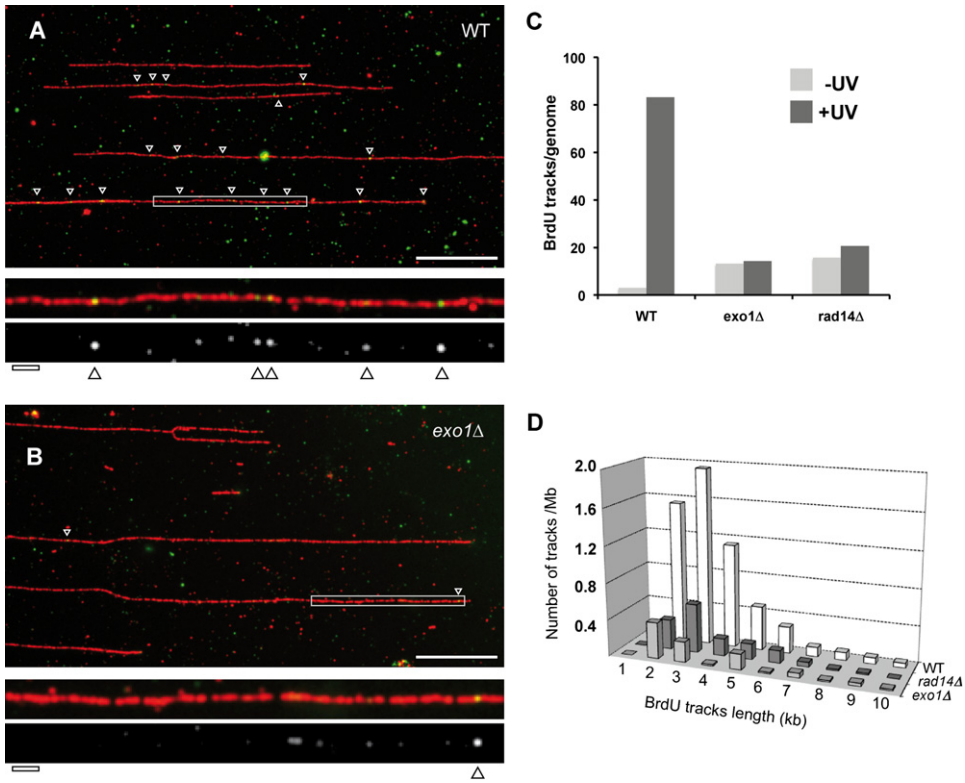


Figure 3. After UV Irradiation, Noncycling Yeast Cells Accumulate Rad14- and Exo1-Dependent BrdU Chromosomal Tracks

WT, *exo1*Δ, and *rad14*Δ strains in a *rho*° *bar1*Δ *cdc45*^{td} background were arrested in G1, UV irradiated (75 J/m^2), and labeled with BrdU as in Figure 2C. Cells were fixed with sodium azide, DNA fibers were stretched by DNA combing, and BrdU tracks (marked by white arrowheads) were visualized by immunofluorescence.

(A and B) Representative images of the BrdU tracks on chromosome fibers. Green: BrdU; red: DNA. The bar represents 50 kb (5 kb for the inset).

(C) Quantitative data relative to BrdU tracks.

(D) Size distribution of the BrdU tracks.

BrdU, incorporated as DNA synthesis tracks into chromosomal DNA fibers, was monitored by immunofluorescence with anti-BrdU antibodies. In Figure 3, we show representative images of DNA fibers stained for DNA and for BrdU. In our experimental conditions, UV-irradiated WT cells exhibit several BrdU tracks (Figure 3A). UV treatment causes an almost 30-fold increase in WT cells, while the lack of *RAD14* or *EXO1* prevents their accumulation (Figures 3B and 3C and data not shown); this indicates that the generation of long BrdU tracks requires both NER and Exo1 activities. The size and frequencies of these refilling patches are analyzed in the graph shown in Figure 3D, and Figure S3 reports the actual numbers and statistical analysis. It has to be noted that DNA combing is limited in detecting short (<2 kb) tracks of BrdU so NER patches around 30 nt in length are not detectable by this approach. These data indicate that upon UV irradiation and after NER starts acting on UV lesions, Exo1 activity can further process some repair intermediates to generate longer ssDNA gaps that are then refilled by long-patch repair synthesis.

Electron Microscopy Provides the Physical Proof that Exo1 Generates ssDNA Gaps on UV-Damaged Chromosomes

To obtain a direct physical proof for the formation of UV-induced and Exo1-dependent ssDNA and to increase the resolution toward shorter gaps, we analyzed genomic DNA by transmission electron microscopy (EM). G1-arrested cells were irradiated with UV light, psoralen cross linked to stabilize DNA intermediates, and processed for the preparation of genomic DNA (see Supplemental Experimental Procedures for details).

EM analysis clearly revealed that DNA molecules derived from UV-irradiated WT cells contained far more ssDNA gaps than the background level observed in untreated cells (Figures 4A–4C, actual numbers are reported in Figure S5A). Intriguingly, no UV-dependent induction of ssDNA gaps could be detected in the absence of Exo1 nuclease or in *rad14Δ* cells, suggesting that such gaps must originate from processing of UV-damaged chromosomes initiated by NER and followed by Exo1 activity (Figure 4C). The gapped molecules visualized by EM are coincident with the intermediates we identified by PFGE analysis. In fact, Figure 4D shows that similarly gapped molecules can be observed when DNA is directly extracted from the cloud in the PFGE itself (Figure S4B).

A kinetics analysis, estimating the number and size of the ssDNA gaps at later time points after UV irradiation of WT cells, is shown in Figure 4E, where the number of ssDNA gaps above background level is plotted as the fold change with respect to gaps detected in untreated cells. While the number of smaller

Please cite this article in press as: Giannattasio et al., Exo1 Competes with Repair Synthesis, Converts NER Intermediates to Long ssDNA Gaps, and Promotes Checkpoint Activation, Molecular Cell (2010), doi:10.1016/j.molcel.2010.09.004

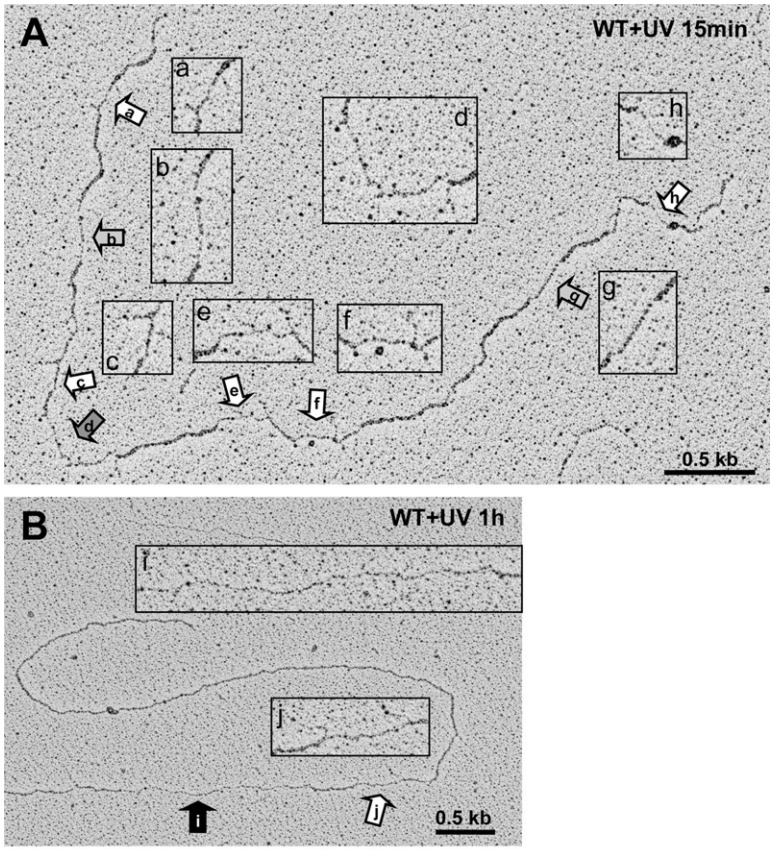


Figure 4. EM Reveals that Rad14- and Exo1-Dependent ssDNA Gaps Accumulate on Chromosomes after UV Irradiation

WT, *exo1*Δ, and *rad14Δ* cells, in a YMG975 background, were arrested in G1 and UV irradiated as in Figure 3. ssDNA gaps were analyzed on DNA fibers 15, 60, and 120 min after UV treatment.

(A and B) Representative pictures of the ssDNA gaps visualized 15 min and 60 min after the treatment. Insets exhibit a two-fold enlargement of the selected regions.

(C) Size distribution of the ssDNA gaps from the 15 min samples.

(D) Agarose plugs from the experiment described in (A) were prepared 15 min after UV irradiation and separated by PFGE. Chromosome intermediates migrating as a cloud were extracted from the PFGE (Figure S4B) and visualized by EM. The graph represents the size distribution of the ssDNA gaps.

(E) Size distribution of the ssDNA gaps in WT cells during the time-course experiment described in (A). The graph reports the fold increase over untreated samples for each gap size class.

ssDNA gaps (<100 nt) reverts close to background levels within 1 hr (probably because of refilling events), larger ssDNA gaps tend to persist for a longer time: they are stabilized well above background levels even 2 hr after UV treatment. Importantly, we do not detect significant time-dependent accumulation of extended gaps in *exo1Δ* cells (Figure S5B). These data provide direct evidence that in WT yeast cells Exo1 can process NER intermediates, giving rise to long ssDNA gaps, which are refilled during DNA repair.

UV-Induced DNA-Damage Checkpoint Is Triggered by Problematic Refilling of the ssDNA Gaps during Repair

Because Exo1-dependent extension of NER gaps is unlikely to happen at all lesion sites, we hypothesized that Exo1 may compete for gap processing with the refilling polymerase during repair and succeed in generating extended ssDNA regions when completion of repair is somehow impaired. This could be related to a defective repair synthesis due to limiting factors or to damaged templates, which may stall the refilling DNA polymerase.

If this hypothesis is correct, we expected that by limiting some repair synthesis factor we should increase the probability of generating extended ssDNA regions and trigger a checkpoint response with fewer lesions. *pcna-cs* cells, which at restrictive temperature cannot synthesize DNA, were arrested in G1 at the permissive temperature and then shifted to lower temperature to inactivate proliferating cell nuclear antigen (PCNA). Cultures were irradiated with different UV doses, and Rad53 phosphorylation and activity were followed in WT and *pcna-cs* strains held in G1. Although in these conditions WT cells start to exhibit phosphorylated Rad53 at 20 J/m^2, a *pcna-cs* mutant clearly activates the checkpoint at 5 J/m^2 or less (Figure 5A), supporting the hypothesis. The strong checkpoint activation detected in *pcna-cs* cells correlates with the accumulation of ssDNA in the PFGE cloud and depends upon Exo1 (Figures 5B and 5C).

Blocking repair DNA synthesis with the Ara-C inhibitor, which prevents elongation of the DNA chain, also provokes a stronger Exo1-dependent checkpoint activation (Figure 5D).

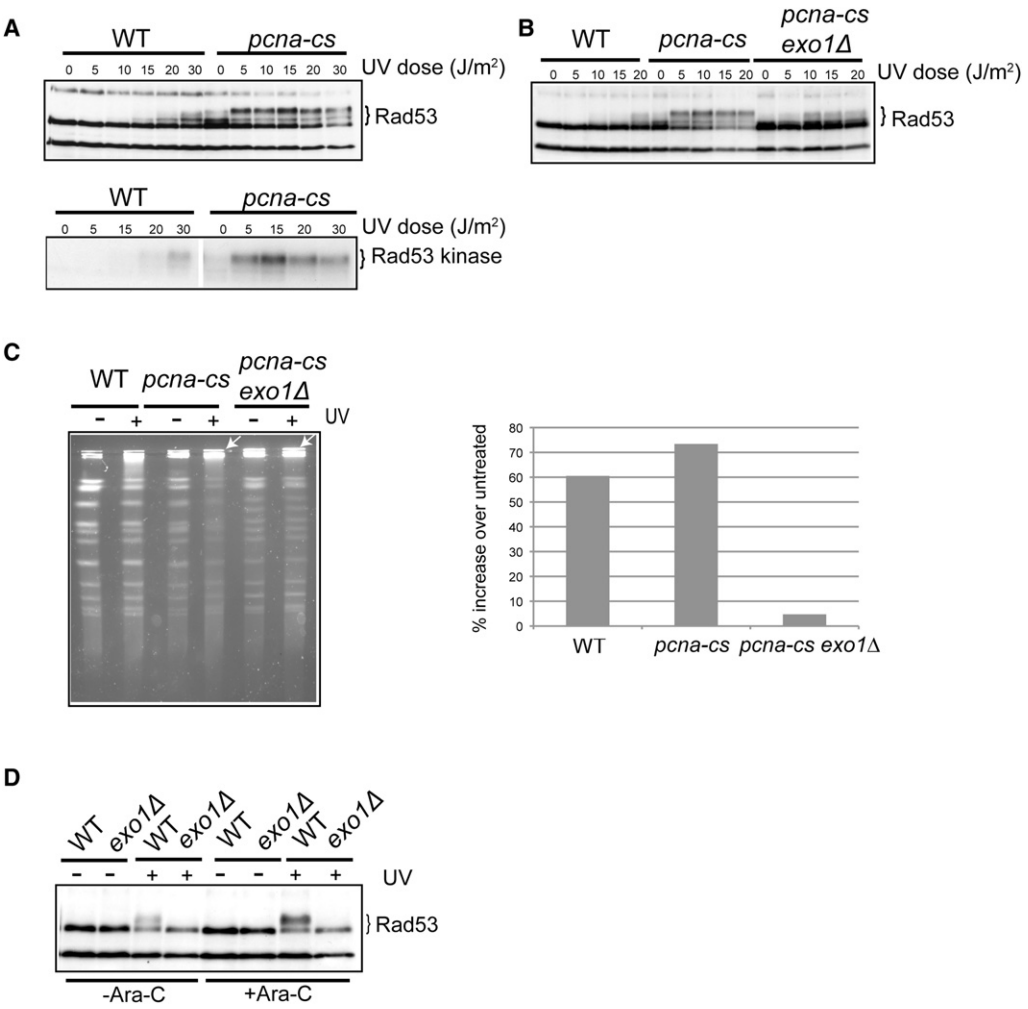


Figure 5. Blockage of DNA Repair Synthesis Allows a Strong Checkpoint Response Even at Very Low UV Doses

pcna44-52 and the corresponding WT cells were arrested in G1, shifted to 12°C, and held in these conditions for 8 hr, before they were treated with the indicated UV doses. Rad53 phosphorylation was detected by western blotting, and Rad53 activity was monitored by in situ kinase assay (A).

(B) WT, *pcna-cs*, and *pcna-cs exo1Δ* cells were treated as in (A) and Rad53 phosphorylation was monitored by western blotting.

(C) The same strains as in (B) have been arrested in G1 and UV treated (75 J/m^2). Chromosome intermediates were detected as in Figure 2. The cloud signals before and after UV treatment have been quantified and normalized to the total amount of DNA loaded.

(D) WT and *exo1Δ* cells (YMG975 background) were arrested in G1, as in Figure 2C, and treated with Ara-C (400 μg/ml) or left untreated. Cells were then UV irradiated (20 J/m^2) and kept arrested 5 min in G1 at 37°C with or without Ara-C. Rad53 phosphorylation was detected by western blotting.

Closely Opposing DNA Lesions Could Be Involved in Activating the Response to UV Irradiation

During these studies, we observed that Rad53 activation, reflecting checkpoint response, increases with the square of the UV dose (Figure S6A). This suggests that checkpoint activation may entail a two-hits mechanism and that closely opposing lesions may be involved: processing of one lesion by NER would leave the opposing lesion in the template strand during repair synthesis, blocking the DNA polymerase. In order to complete the refilling reaction in G1, cells would have to employ translesion DNA polymerases. Elimination of TLS polymerases should increase the accumulation of blocked intermediates, which can be processed by Exo1, resulting in stronger checkpoint activation. *TLSΔ* cells, lacking the genes coding for all known yeast translesion polymerases, were held in G1 and irradiated with different UV doses. Figure 6A shows that Rad53 phosphorylation is greatly increased in *TLSΔ* cells compared to WT cells, implying that TLS polymerases, bypassing template strand lesions, can quench checkpoint signaling. Checkpoint hyperactivation in these conditions still requires *EXO1* (Figure 6B) and correlates to the *EXO1*-dependent accumulation of ssDNA intermediates in PFGE (Figure 6C). Moreover, epistasis analysis indicates that TLS polymerases function downstream of NER (Figure S6B). In an effort to identify the DNA polymerase implicated in this process, we analyzed individual deletions in *RAD30*, *REV1*, and *REV3*. Our results indicate that although the complete loss of TLS activities has the stronger effect, each of the single mutants exhibits an increased checkpoint signaling

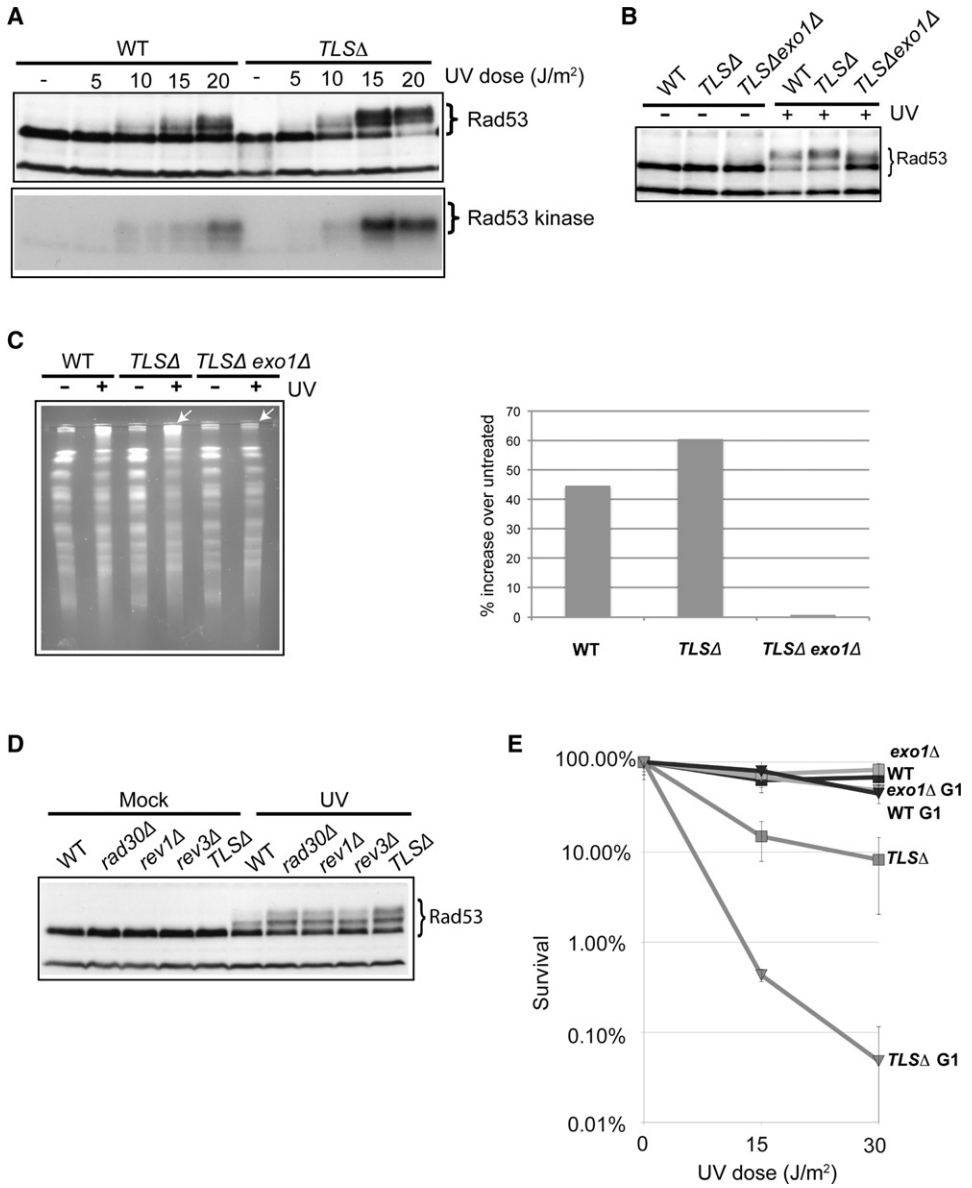


Figure 6. Rad53 Activity Is Enhanced in the Absence of TLS Polymerases

(A) WT and *TLSΔ* (YMG1082 *rev1*Δ, *rev3*Δ, *rev7*Δ, and *rad30*Δ) cells were G1 arrested and treated with the indicated UV doses. Rad53 phosphorylation and kinase activity were detected 30 min after the UV treatment by western blotting and in situ kinase assay, respectively.

(B) WT, *TLSΔ* (YMG1082), and *TLSΔexo1*Δ cells were arrested in G1 and treated with 75 J/m^2 UV. Rad53 phosphorylation was detected by western blotting.

(C) The indicated strains (*TLSΔ* = *rev1*Δ, *rev3*Δ, *rev7*Δ, *rad30*Δ) were arrested and treated as in (B). Chromosome intermediates were detected by PFGE and quantified as in Figure 5C.

(D) The indicated strains (*TLSΔ* = *rev1*Δ, *rev3*Δ, *rev7*Δ, *rad30*Δ) were arrested in G1 and treated with 15 J/m^2 UV. Rad53 phosphorylation was detected by western blotting.

(E) WT, *exo1Δ*, and *TLSΔ* (*TLSΔ* = *rev1*Δ, *rev3*Δ, *rev7*Δ, *rad30*Δ) cells were arrested in G1 or kept in cycling conditions. Cultures were treated with the indicated UV doses and plated on rich medium. The percentage of surviving cells at different UV dosages was scored after 3 days incubation. Error bars represent standard deviation calculated from three independent experiments.

(Figure 6D). Moreover, while WT cells exhibit a time-dependent dephosphorylation of Rad53, hyperphosphorylated Rad53 is still present at high levels 2 hr after UV irradiation in *TLSΔ* cells (Figure S6C). Interestingly, *TLSΔ* cells UV irradiated in G1 and released into the cell cycle fail to proceed into S phase in a timely manner, possibly because in G1 the refilling process cannot be completed in the absence of TLS polymerases, gapped molecules cannot be effectively replicated, and the checkpoint is not switched off (Figure S6D). If indeed this mechanism is active due to the presence of closely opposing lesions, we would

expect TLS activity to be extremely important in UV-irradiated G1 cells, where translesion is the only way to allow completion of repair synthesis. We compared the sensitivity of WT and *TLSΔ* cells treated with UV light in G1-arrested or in unsynchronized cultures. Figure 6E shows that *TLSΔ* cells are extremely sensitive when UV irradiated in G1, while for WT or *exo1Δ* strains the cell-cycle phase in which the cells are hit with UV light does not seem to be particularly relevant, indicating a crucial role for TLS polymerases in helping noncycling cells survive UV irradiation.

DISCUSSION

This study provides physical evidence that Exo1 processes NER intermediates when repair synthesis is impeded: this event may be caused by a low level in DNA synthesis factors, by low levels of dNTPs, or by the presence of closely opposing lesions. Exo1 converts NER intermediates to long ssDNA gaps, which then promote DNA-damage checkpoint activation in noncycling cells and channel the problematic lesion to different repair mechanisms.

Survival after UV irradiation is warranted by NER, which is the only system repairing UV lesions in humans. In yeast cells arrested in G1 or G2 and in human resting fibroblasts, the immediate response to UV irradiation is lost in the absence of NER (Giannattasio et al., 2004; Marini et al., 2006; Marti et al., 2006), suggesting that NER intermediates may be required to activate a prompt DNA-damage checkpoint. Here we analyze the mechanism in charge of the rapid response to UV irradiation in noncycling cells and identify the molecular determinants for the cellular response to UV light.

ssDNA is involved in the DSBs response and ssDNA gaps are one of the putative intermediates of NER. However, NER-mediated excision of damaged DNA oligonucleotides is so tightly coordinated with actual refilling that ssDNA regions are very short-lived (Staresincic et al., 2009). Moreover, ssDNA gaps generated during NER are expected to be very short and inefficient in activating the checkpoint. We hypothesized that nuclease activities may further process some NER intermediates and may thus be involved in the checkpoint response.

The results reported in this work demonstrate that yeast cells lacking *EXO1* are completely defective in achieving a functional G1 checkpoint arrest and Rad53 phosphorylation. In the signaling cascade, Exo1 seems to act at the level of the sensors, since two direct substrates of Mec1, Ddc2 and H2A, are not targeted by the kinase in *exo1Δ* cells. The requirement for Exo1 in generating a checkpoint signal probably impinges on the production of enough ssDNA to allow recruitment and activation of checkpoint sensors; indeed, mutations in the catalytic site of Exo1 recapitulate the phenotypes exhibited by the complete loss of *EXO1*. UV light induces the accumulation of DNA structures running in PFGE as a cloud close to the wells of the gel. These structures closely correlate with checkpoint activation and are virtually lost in *exo1Δ* and in *rad14Δ* cells, strongly suggesting that once NER has started to repair the lesions, Exo1 processes UV-damaged chromosomes generating ssDNA gaps. This conclusion is supported by the findings that the cloud intermediates are sensitive to the ssDNA-specific S1 nuclease, and BrdU incorporation (a direct marker of ongoing repair DNA synthesis) in the cloud was entirely dependent upon NER and largely dependent on Exo1 activity.

Since NER activity is not expected to give rise to long ssDNA gaps, it was important to obtain physical evidence for the formation of such ssDNA regions. We followed two approaches: DNA combing and EM.

Measuring indirectly ssDNA gaps by allowing DNA repair synthesis to occur in the presence of BrdU and monitoring its incorporation into repair tracks by DNA combing, we show the formation of long UV-induced BrdU tracks within chromosome fibers, suggesting that, during NER, some lesions may be repaired through a mechanism recalling the long-patch NER described several years ago in bacteria (Cooper, 1982). Strikingly, formation of these long tracks requires NER activity and Exo1 function. By EM, we prove that UV irradiation of WT G1 cells results in the accumulation of molecules containing ssDNA gaps of various size, which depend upon UV treatment and a functional Exo1 nuclease. Admittedly, there is also a background level of DNA discontinuities in untreated samples, which is known to be associated with the technical procedure. For this reason, we cannot distinguish between the nicks due to the procedure from the NER-specific 30 nt gaps and those from extended ssDNA gaps that are between 30 and 100 nt long. What is nonetheless clear from the statistical analysis is that, in UV-damaged chromosomes, Exo1 mediates the formation of ssDNA gaps, many of which are much longer than the average NER-derived gap. We propose that gapped molecules probably originate, through a dynamic process, from a competition between Exo1 exonucleolytic activity and the refilling activity of DNA polymerases. Accordingly, we do see a time-dependent increase in the accumulation of long ssDNA gaps, while shorter ones are progressively decreasing, possibly because of refilling, leading at later time points to a steady-state accumulation of larger ssDNA gaps above background levels. Using EM analysis, we hardly detected gaps longer than 1 kb. Two non-mutually exclusive explanations can be foreseen: long ssDNA regions are intrinsically more prone to breakage during sample preparation; or, since EM provides an instant picture of a dynamic process (ssDNA generation and gap refilling), very long ssDNA gaps may be indeed rare in vivo. It is hard to avoid noticing a similarity between our observations and those of a pioneering paper that described bacterial long-patch excision repair of UV lesions (Cooper, 1982). Our data are in agreement with previous findings reporting that NER activity is required to achieve UV-induced ubiquitination of histone H2A in mammalian cells, and such modification is enhanced by interfering with repair DNA synthesis (Marteijn et al., 2009).

We tried to determine why some lesions are extensively processed by Exo1 instead of being simply repaired by NER. If repair synthesis and Exo1 compete for the gap, and the refilling reaction at the end of NER is somehow impeded, the resulting structure could resemble that of a stalled replicating polymerase. This would leave a 5′-ended filament that could be digested by Exo1, extending the gap; this view is supported by previous evidence obtained in human cells (Matsumoto et al., 2007). This hypothesis can be tested by blocking the refilling reaction

Please cite this article in press as: Giannattasio et al., Exo1 Competes with Repair Synthesis, Converts NER Intermediates to Long ssDNA Gaps, and Promotes Checkpoint Activation, Molecular Cell (2010), doi:10.1016/j.molcel.2010.09.004

at most repair sites, which should lead to a higher sensitivity of the checkpoint response and a greater Rad53 activation at lower UV doses. Indeed, when we genetically (with a *pcna-cs* mutant) or chemically (with the Ara-C DNA polymerase inhibitor) impeded refilling, we observed a much stronger signaling. Problems during gap filling can arise when DNA synthesis factors become limiting, as we observed employing a *pcna-cs* mutant, or if the DNA polymerase is blocked, for example when a UV lesion is found on the template strand. Intriguingly, we found that checkpoint activity, measured by Rad53 kinase assays, increases approximately with the square of the UV dose, suggesting that, in our conditions, two lesions may need to concur for the efficient generation of the checkpoint signal. Such events, known as closely opposing lesions, although being statistically unlikely, are known to happen and at 100 J/m^2 represent approximately 1% of all lesions (Lam and Reynolds, 1986, 1987; Sedgwick, 1976; Friedberg et al., 2006; Svetlova et al., 2002). Moreover, pioneering work suggested that in bacteria, a strong SOS-inducing signal in *uvr*$^+$-irradiated cells may be ascribed to the presence of a lesion opposite an excision-repair-formed gap (Salles and Defais, 1984; Bridges and Brown, 1992). In such a system, extensive evidence suggests that the SOS response is not directly elicited by the presence of UV-induced lesions, but the signal is produced when cells attempt to replicate damaged DNA (Salles and Defais, 1984). In nonreplicating cells, SOS induction was detected only in excision-repair-proficient cells, while it was undetectable in repair-deficient cells. These findings suggested that, in the absence of replication, SOS induction may originate from the removal of lesions and the appearance of DNA gaps (Salles and Defais, 1984; Friedberg et al., 2006).

NER cannot remove two closely opposing lesions at the same time (Svoboda et al., 1993). When the lesion on one filament is processed by NER, the refilling polymerase would hit the second lesion on the template and stall; this would tilt the balance between refilling and nucleolytic processing in favor of Exo1, leading to the generation of long ssDNA gaps. Such a model predicts that TLS polymerase activity may be involved in refilling these gaps to overcome the blocking lesion. This is supported by our results showing that in the absence of TLS polymerases, checkpoint signaling is much stronger and TLS mutant cells are extremely sensitive to UV light when irradiated in G1. In fact, in the absence of TLS, haploid cells in G1 have no other means to refill a gap containing a lesion in the template strand. A complementary explanation of these findings rests on recent reports demonstrating that mammalian pol κ is directly involved in NER, at least at some lesion sites (Ogi and Lehmann, 2006; Ogi et al., 2010). While budding yeast lacks pol κ and a similar function for other yeast TLS polymerases has not been reported yet, it is possible that this mechanism will prove to be conserved across evolution and may contribute to the effects we report in TLS mutant cells.

Much of the evidence we presented here supports the view that the UV-induced checkpoint depends upon the formation of long ssDNA gaps. We cannot formally exclude the possibility that very few gaps may fuse and/or break during processing, but Figure S7 supports our conclusion. One single chromosomal break is lethal for *rad52Δ* cells (Vaze et al., 2002); on the other hand, *rad52Δ* are not particularly sensitive to UV light, suggesting that no DSB are formed.

The model we present (Figure 7) is derived from our results with UV light and NER, but it may be extended to other kinds of DNA lesions and repair mechanisms. We propose that during repair of UV lesions DNA polymerases attempt refilling the gaps, while Exo1 tries to extend ssDNA regions. Normally, since DNA synthesis proceeds ~20 times faster than Exo1 (3700 bp/min versus 160 bp/min; Morin et al., 2008), the gap is rapidly refilled. When something impedes DNA synthesis, Exo1 has the chance to generate extended ssDNA gaps that trigger checkpoint activation. We also propose that closely opposing lesions may be involved in such a mechanism and that a repair DNA polymerase blocked by a template lesion represents a structure closely similar to a blocked replicating polymerase, after repriming has taken place 3′ to the blocking lesion. Exposure to elevated UV doses increases the probability of generating closely opposing lesions; upon NER processing of these lesions in G1-arrested haploid cells, a TLS event would be required to complete NER DNA synthesis. Importantly, our data uncover and describe the molecular mechanism of the essential role played by TLS polymerases in UV-irradiated noncycling cells and an unexpected checkpoint-quenching activity of translesion synthesis. These data complement previous reports on pol κ function in NER (Ogi and Lehmann, 2006; Ogi et al., 2010). Finally, the conversion of problematic NER intermediates to long ssDNA gaps may be crucial to channel the intermediates to a different repair and/or bypass mechanism, for example, UV-induced recombination that is promoted by extended ssDNA gaps (Mozlin et al., 2008). In agreement with this view, a considerable body of evidence suggests that NER influences the timing and mechanism of UV-induced mutagenesis. In yeast, mutations due to UV light are TLS dependent and arise mostly in the prereplicative phase in NER-proficient cells, while in NER-deficient cells mutations are fixed postreplicatively (Eckardt et al., 1980). A similar situation has been described for *Escherichia coli* (Bridges and Mottershead, 1971), where a working model explaining these findings is based on processing of closely opposing lesions. Consistently, Exo1 has been suggested to be involved in a MMR-independent mutation avoidance pathway (Tran et al., 2002, 2004, 2007). It will be important to follow up on these studies and determine the effect of Exo1 mutations on long-term events, such as UV-induced mutagenic load and chromosome stability in cycling cells.

In replicating cells, failure to activate a G1 checkpoint after DNA damage will allow lesions persistence in S phase, leading to genome instability. This is particularly important in cells where the apoptotic pathway is misfunctional; intriguingly, Exo1 is required both to properly activate the G1 checkpoint and to promote apoptosis (Bolderson et al., 2009), suggesting that Exo1 activity may be crucial to prevent tumor development induced by DNA lesions.

EXPERIMENTAL PROCEDURES

Strains

All of the strains used in this work are derivatives of W303 (*MAT*a *ade2-1 trp1-1 can1-100 leu2-3,12 his3-11,15 ura3 rad5-535*) and are listed in Table S1.

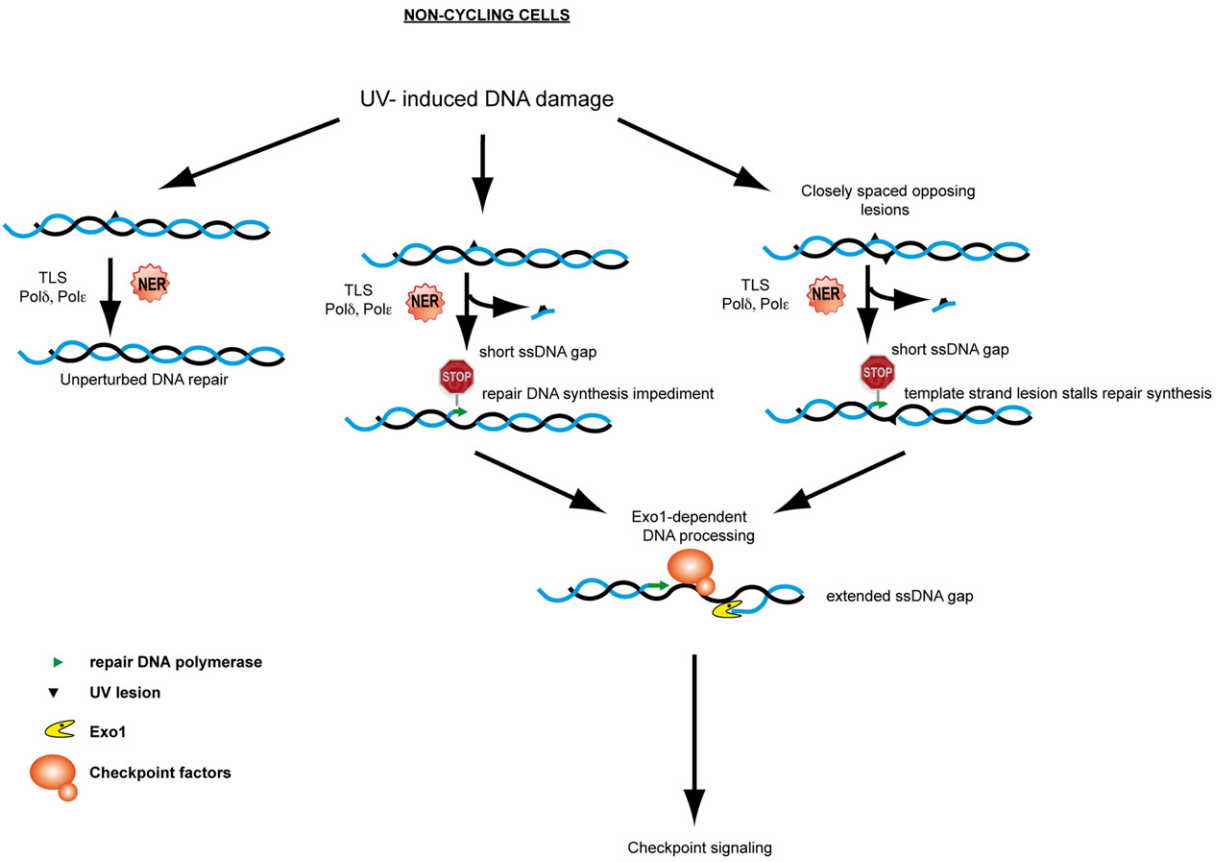


Figure 7. Model for the Checkpoint Activation after UV Irradiation

In noncycling cells NER starts to process UV lesions. Refilling of the NER gap competes with gap extension by Exo1. Impediments to repair DNA synthesis, (e.g., low levels of DNA synthesis factors or DNA polymerase encountering a second lesion on the template strand) may tilt the competition in favor of Exo1, which can further process the intermediate extending the ssDNA gap. Large ssDNA gaps can recruit checkpoint complexes, leading to checkpoint activation. The activity of TLS polymerases promotes gap filling at some lesions sites (e.g., closely opposing lesions or particular sites), in competition with Exo1-dependent gap extension, and counteracts checkpoint activation.

Yeast mutants (deleted for nuclease-encoding genes *APN1*, *APN2*, *SAE2*, *MRE11*, *MKT1*, *RAD27*, *YEN1*, *DIN7*, and *DNA2*) were also analyzed for possible checkpoint defects and tested negative.

UV Irradiation

G1-arrested cells were irradiated on plates with 75 J/m^2, unless otherwise indicated, collected, and processed as described (Giannattasio et al., 2004). Protein extracts, SDS-PAGE conditions, checkpoint assays, and Fluorescence-activated cell sorter (FACS) analysis have been performed as previously described (Giannattasio et al., 2004). Rad53 in situ kinase assays were performed as described (Pellicioli et al., 1999).

PFGE, BrdU Incorporation Detection, Southern-Blotting Analysis, S1 Assay, and DNA Combing

α-factor-arrested cells were UV irradiated and held in G1 for 2 hr to allow repair in the presence of BrdU. Samples were then analyzed by DNA combing (Supplemental Experimental Procedures). To avoid BrdU incorporation from chromosomal and mitochondrial DNA replication, we used rho° and *bar1*Δ strains that express a Cdc45 degron fusion protein obtained with the one-step system (Tercero et al., 2000) and the experimental conditions described in Supplemental Experimental Procedures.

PFGE was performed with an Amersham Gene Navigator system (Supplemental Experimental Procedures). BrdU incorporation in PFGE was detected as described (Lengronne et al., 2001). Southern-blot analysis was performed in standard conditions with labeled PCR fragments (Supplemental Experimental Procedures).

Electron Microscopy

Cells were UV irradiated and, at different times after treatment, were placed on ice with 1% sodium azide for 1 hr and cross linked with psoralen as described (Lopes, 2009). Genomic DNA was prepared and digested with PvuI before EM analysis. For details see the Supplemental Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and one table and can be found with this article online at doi:10.1016/j.molcel.2010.09.004.

ACKNOWLEDGMENTS

We thank E. Schwob and the DNA combing facility of Montpellier for silanized coverslips. We are grateful to P. Burgers and M. Liskay for providing critical mutant strains and to M. Foiani and D. Branzei for critically reading the manuscript. All the members of the laboratory and A. Lehmann are thanked for stimulating discussions. The laboratory of P. Pasero is supported by Fondation

Recherche Mèdicale, Agence Nationale de la Recherche, and Institut National du Cancer. M.L. and C.F. are supported by Swiss National Science Foundation Grant PP00A-114922. This work was supported by grants from AIRC and Fondazione Cariplo to P. Plevani and M.M.-F and the European Union FP6 Integrated Project DNA Repair contract 512113 to P. Plevani. The financial support of Telethon–Italy (grant GGP030406 to M.M.-F.) is gratefully acknowledged.

Received: February 23, 2010
Revised: May 26, 2010
Accepted: July 23, 2010
Published: October 7, 2010

REFERENCES

Bolderson, E., Richard, D.J., Edelmann, W., and Khanna, K.K. (2009). Involvement of Exo1b in DNA damage-induced apoptosis. Nucleic Acids Res. *37*, 3452–3463.

Bridges, B.A., and Brown, G.M. (1992). Mutagenic DNA repair in Escherichia coli. XXI. A stable SOS-inducing signal persisting after excision repair of ultraviolet damage. Mutat. Res. *270*, 135–144.

Bridges, B.A., and Mottershead, R. (1971). RecA + -dependent mutagenesis occurring before DNA replication in UV- and -irradiated Escherichia coli. Mutat. Res. *13*, 1–8.

Chen, Z., Graham, R., Burns, M.A., and Larson, R.G. (2007). Modeling ssDNA electrophoretic migration with band broadening in an entangled or cross-linked network. Electrophoresis *28*, 2783–2800.

Cooper, P.K. (1982). Characterization of long patch excision repair of DNA in ultraviolet-irradiated Escherichia coli: an inducible function under rec-lex control. Mol. Gen. Genet. *185*, 189–197.

Cordeiro-Stone, M., Schumacher, R.I., and Meneghini, R. (1979). Structure of the replication fork in ultraviolet light-irradiated human cells. Biophys. J. *27*, 287–300.

Cotta-Ramusino, C., Fachinetti, D., Lucca, C., Doksani, Y., Lopes, M., Sogo, J., and Foiani, M. (2005). Exo1 processes stalled replication forks and counteracts fork reversal in checkpoint-defective cells. Mol. Cell *17*, 153–159.

Eckardt, F., Teh, S.J., and Haynes, R.H. (1980). Heteroduplex repair as an intermediate step of UV mutagenesis in yeast. Genetics *95*, 63–80.

Friedberg, E.C., Walker, G.C., and Siede, W. (2006). DNA Repair and Mutagenesis (Washington, D.C.: ASM Press).

Giannattasio, M., Lazzaro, F., Longhese, M.P., Plevani, P., and Muzi-Falconi, M. (2004). Physical and functional interactions between nucleotide excision repair and DNA damage checkpoint. EMBO J. *23*, 429–438.

Goodman, M.F. (2002). Error-prone repair DNA polymerases in prokaryotes and eukaryotes. Annu. Rev. Biochem. *71*, 17–50.

Harrison, J.C., and Haber, J.E. (2006). Surviving the breakup: the DNA damage checkpoint. Annu. Rev. Genet. *40*, 209–235.

Kerzendorfer, C., and O'Driscoll, M. (2009). Human DNA damage response and repair deficiency syndromes: linking genomic instability and cell cycle checkpoint proficiency. DNA Repair (Amst.) *8*, 1139–1152.

Lam, L.H., and Reynolds, R.J. (1986). Repair of closely opposed cyclobutyl pyrimidine dimers in UV-sensitive human diploid fibroblasts. Mutat. Res. *166*, 199–205.

Lam, L.H., and Reynolds, R.J. (1987). DNA sequence dependence of closely opposed cyclobutyl pyrimidine dimers induced by UV radiation. Mutat. Res. *178*, 167–176.

Lazzaro, F., Giannattasio, M., Puddu, F., Granata, M., Pellicioli, A., Plevani, P., and Muzi-Falconi, M. (2009). Checkpoint mechanisms at the intersection between DNA damage and repair. DNA Repair (Amst.) *8*, 1055–1067.

Lehmann, A.R. (1979). The relationship between pyrimidine dimers and replicating DNA in UV-irradiated human fibroblasts. Nucleic Acids Res. *7*, 1901–1912.

Lehmann, A.R., and Fuchs, R.P. (2006). Gaps and forks in DNA replication: rediscovering old models. DNA Repair (Amst.) *5*, 1495–1498.

Lengronne, A., Pasero, P., Bensimon, A., and Schwob, E. (2001). Monitoring S phase progression globally and locally using BrdU incorporation in TK(+) yeast strains. Nucleic Acids Res. *29*, 1433–1442.

Lopes, M. (2009). Electron microscopy methods for studying in vivo DNA replication intermediates. Methods Mol. Biol. *521*, 605–631.

Lopes, M., Foiani, M., and Sogo, J.M. (2006). Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. Mol. Cell *21*, 15–27.

Marini, F., Nardo, T., Giannattasio, M., Minuzzo, M., Stefanini, M., Plevani, P., and Muzi Falconi, M. (2006). DNA nucleotide excision repair-dependent signaling to checkpoint activation. Proc. Natl. Acad. Sci. USA *103*, 17325–17330.

Marteijn, J.A., Bekker-Jensen, S., Mailand, N., Lans, H., Schwertman, P., Gourdin, A.M., Dantuma, N.P., Lukas, J., and Vermeulen, W. (2009). Nucleotide excision repair-induced H2A ubiquitination is dependent on MDC1 and RNF8 and reveals a universal DNA damage response. J. Cell Biol. *186*, 835–847.

Marti, T.M., Hefner, E., Feeney, L., Natale, V., and Cleaver, J.E. (2006). H2AX phosphorylation within the G1 phase after UV irradiation depends on nucleotide excision repair and not DNA double-strand breaks. Proc. Natl. Acad. Sci. USA *103*, 9891–9896.

Matsumoto, M., Yaginuma, K., Igarashi, A., Imura, M., Hasegawa, M., Iwabuchi, K., Date, T., Mori, T., Ishizaki, K., Yamashita, K., et al. (2007). Perturbed gap-filling synthesis in nucleotide excision repair causes histone H2AX phosphorylation in human quiescent cells. J. Cell Sci. *120*, 1104–1112.

Morin, I., Ngo, H.P., Greenall, A., Zubko, M.K., Morrice, N., and Lydall, D. (2008). Checkpoint-dependent phosphorylation of Exo1 modulates the DNA damage response. EMBO J. *27*, 2400–2410.

Mozlin, A.M., Fung, C.W., and Symington, L.S. (2008). Role of the Saccharomyces cerevisiae Rad51 paralogs in sister chromatid recombination. Genetics *178*, 113–126.

Nakada, D., Hirano, Y., and Sugimoto, K. (2004). Requirement of the Mre11 complex and exonuclease 1 for activation of the Mec1 signaling pathway. Mol. Cell. Biol. *24*, 10016–10025.

Ogi, T., and Lehmann, A.R. (2006). The Y-family DNA polymerase kappa (pol kappa) functions in mammalian nucleotide-excision repair. Nat. Cell Biol. *8*, 640–642.

Ogi, T., Limsirichaikul, S., Overmeer, R.M., Volker, M., Takenaka, K., Cloney, R., Nakazawa, Y., Niimi, A., Miki, Y., Jaspers, N.G., et al. (2010). Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. Mol. Cell *37*, 714–727.

Pellicioli, A., Lucca, C., Liberi, G., Marini, F., Lopes, M., Plevani, P., Romano, A., Di Fiore, P.P., and Foiani, M. (1999). Activation of Rad53 kinase in response to DNA damage and its effect in modulating phosphorylation of the lagging strand DNA polymerase. EMBO J. *18*, 6561–6572.

Rattray, A.J., and Strathern, J.N. (2003). Error-prone DNA polymerases: when making a mistake is the only way to get ahead. Annu. Rev. Genet. *37*, 31–66.

Rupp, W.D., and Howard-Flanders, P. (1968). Discontinuities in the DNA synthesized in an excision-defective strain of Escherichia coli following ultraviolet irradiation. J. Mol. Biol. *31*, 291–304.

Rupp, W.D., Wilde, C.E., 3rd, Reno, D.L., and Howard-Flanders, P. (1971). Exchanges between DNA strands in ultraviolet-irradiated Escherichia coli. J. Mol. Biol. *61*, 25–44.

Salles, B., and Defais, M. (1984). Signal of induction of recA protein in E. coli. Mutat. Res. *131*, 53–59.

Sedgwick, S.G. (1976). Misrepair of overlapping daughter strand gaps as a possible mechanism for UV induced mutagenesis in UVR strains of Escherichia coli: a general model for induced mutagenesis by misrepair (SOS repair) of closely spaced DNA lesions. Mutat. Res. *41*, 185–200.

Segurado, M., and Diffley, J.F. (2008). Separate roles for the DNA damage checkpoint protein kinases in stabilizing DNA replication forks. Genes Dev. *22*, 1816–1827.

Staresincic, L., Fagbemi, A.F., Enzlin, J.H., Gourdin, A.M., Wijgers, N., Dunand-Sauthier, I., Giglia-Mari, G., Clarkson, S.G., Vermeulen, W., and Schärer, O.D. (2009). Coordination of dual incision and repair synthesis in human nucleotide excision repair. EMBO J. *28*, 1111–1120.

Svetlova, M., Solovjeva, L., Pleskach, N., Yartseva, N., Yakovleva, T., Tomilin, N., and Hanawalt, P. (2002). Clustered sites of DNA repair synthesis during early nucleotide excision repair in ultraviolet light-irradiated quiescent human fibroblasts. Exp. Cell Res. *276*, 284–295.

Svoboda, D.L., Smith, C.A., Taylor, J.S.A., and Sancar, A. (1993). Effect of sequence, adduct type, and opposing lesions on the binding and repair of ultraviolet photodamage by DNA photolyase and (A)BC excinuclease. J. Biol. Chem. *268*, 10694–10700.

Tercero, J.A., Labib, K., and Diffley, J.F. (2000). DNA synthesis at individual replication forks requires the essential initiation factor Cdc45p. EMBO J. *19*, 2082–2093.

Tran, P.T., Erdeniz, N., Dudley, S., and Liskay, R.M. (2002). Characterization of nuclease-dependent functions of Exo1p in Saccharomyces cerevisiae. DNA Repair (Amst.) *1*, 895–912.

Tran, P.T., Erdeniz, N., Symington, L.S., and Liskay, R.M. (2004). EXO1-A multi-tasking eukaryotic nuclease. DNA Repair (Amst.) *3*, 1549–1559.

Tran, P.T., Fey, J.P., Erdeniz, N., Gellon, L., Boiteux, S., and Liskay, R.M. (2007). A mutation in EXO1 defines separable roles in DNA mismatch repair and post-replication repair. DNA Repair (Amst.) *6*, 1572–1583.

Vaze, M.B., Pellicioli, A., Lee, S.E., Ira, G., Liberi, G., Arbel-Eden, A., Foiani, M., and Haber, J.E. (2002). Recovery from checkpoint-mediated arrest after repair of a double-strand break requires Srs2 helicase. Mol. Cell *10*, 373–385.

Ward, I.M., Minn, K., and Chen, J. (2004). UV-induced ataxia-telangiectasia-mutated and Rad3-related (ATR) activation requires replication stress. J. Biol. Chem. *279*, 9677–9680.

Wei, K., Clark, A.B., Wong, E., Kane, M.F., Mazur, D.J., Parris, T., Kolas, N.K., Russell, R., Hou, H., Jr., Kneitz, B., et al. (2003). Inactivation of Exonuclease 1 in mice results in DNA mismatch repair defects, increased cancer susceptibility, and male and female sterility. Genes Dev. *17*, 603–614.

Zhang, H., and Lawrence, C.W. (2005). The error-free component of the RAD6/RAD18 DNA damage tolerance pathway of budding yeast employs sister-strand recombination. Proc. Natl. Acad. Sci. USA *102*, 15954–15959.

Zou, L., and Elledge, S.J. (2003). Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. Science *300*, 1542–1548.

4.2. Mismatch repair events can trigger PCNA ubiquitylation and recruitment of polymerase-h to chromatin.

Javier Peña-Diaz, Stephanie Felscher, Cindy Follonier, Dennis Castor, Massimo Lopes, Alessandro A. Sartori and Josef Jiricny
Institute of Molecular Cancer Research, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

Abstract

U/G mismatches arising in DNA through spontaneous deamination of cytosine are repaired by base excision repair (BER), a process initiated by excision of the uracil by uracil-DNA glycosylase (UDG). However, U/G mispairs arise also in immunoglobulin (Ig) loci of antigen-activated B-cells during somatic hypermutation (SHM) and class switch recombination (CSR) through the action of activation-induced deaminase (AID). Puzzlingly, U/G processing at Ig loci gives rise to numerous mutations in both C/G and T/A base pairs spread over long DNA tracts. That mutations at C/Gs arise through incomplete repair of uracils was shown in cells from patients and mouse models lacking UDG, but, unexpectedly, lack of uracil processing did not affect mutagenesis at T/As. We now show that U/G processing in human cell extracts in which UDG was inhibited is mediated by the mismatch repair (MMR) system. However, in contrast to canonical MMR, which initiates at pre-existing strand discontinuities and directs the repair process to the discontinuous strand, we observed repair synthesis also in covalently-closed substrates. These events lacked strand directionality and required the endonucleolytic activity of MLH1/PMS2. We also detected MMR-dependent ubiquitylation of PCNA, a post-translational modification reported to recruit error-prone polymerases to DNA damage. We thus postulated that a subset of SHM- and CSR-associated mutations might arise through processing of AID-induced U/G mispairs by non-canonical MMR. In support of this hypothesis, activation of MMR *in vivo* was accompanied not only by PCNA ubiquitylation, but also by recruitment to chromatin of polymerase-h, which was shown, like MMR, to be required for mutagenesis at T/A base pairs during SHM and CSR. This mechanism may contribute to global mutagenesis, as

it is likely to be activated by all types of DNA damage arising outside of S-phase that is addressed by MMR.

My contribution

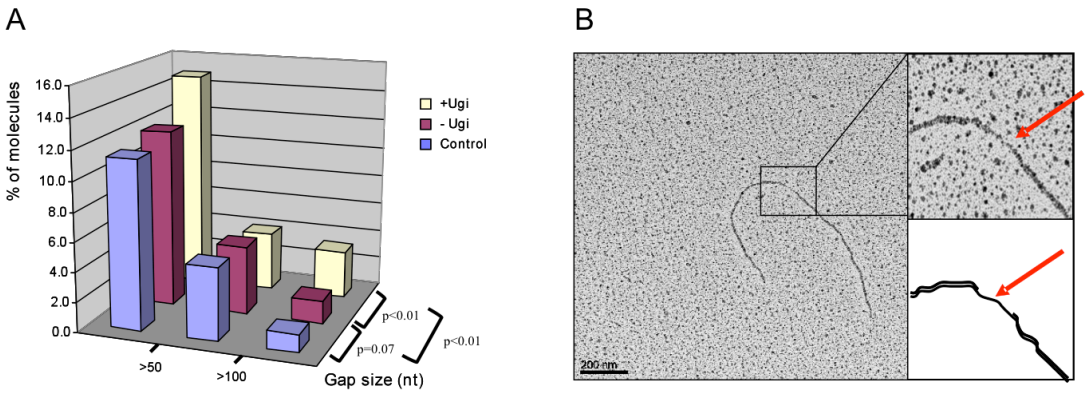


Figure 42. A. Quantification and size distribution of gaps detected in a U/G-U/G substrate recovered after incubation with a BL2 extract. The substrate was linearized. Substrate incubated with heat-inactivated extract was used as control. n>400. B. Electron micrograph of a representative gap. Arrows indicate the ssDNA region in the magnified inset (upper right) and in its graphic representation (lower right). These panels show that BER inhibition leads to an increased number of gapped molecules in the repair reactions.

4.3. Control of DNA replication at budding yeast telomeres

Laure Lemmens and David Shore
Department of Molecular Biology, University of Geneva, Quai Ernest-Ansermet 30, CH-1211 Geneva, Switzerland

Abstract
Several lines of evidence indicate that conventional DNA replication of telomeres is highly coordinated with the action of telomerase, yet the interaction of the replication fork with telomeres is still not well characterized.

Previous studies in both fission and budding yeast have shown that replication fork passage is slowed at or near telomeric repeat tracts. In *S. pombe*, the telomere-repeat binding protein Taz1 plays a direct role in promoting conventional DNA replication of the telomere. Furthermore, several studies have shown that telomeres themselves can exert an effect on the timing of firing of nearby replication origins, or on origin activation, through mechanisms that are still incompletely understood.

We are examining the dynamics of replication fork progression near telomeres in synchronized cell populations, using ChIP to monitor the presence of fork components, in particular Pol2 (the leading-strand Polymerase epsilon) and Pol1 (Polymerase alpha, involved in initiation and lagging-strand synthesis). Our results indicate that Pol2 and Pol1 display identical binding kinetics at both early- and late-firing replication origins, whereas Pol1 binds for a longer period of time than Pol2 when the replication fork arrives at telomeres. One possible model to explain these observations would be that the lagging strand polymerase is strongly slowed down when arriving close to a telomere, compared to Pol2, generating a physical and temporal uncoupling of leading- and lagging-strand synthesis at telomeres.

Additional experiments from us and others show that deletion of the telomeric binding protein Rif1 induces earlier telomere replication compare to wild-type

cells, and abolishes the difference in the kinetics of Pol1 and Pol2 binding at telomere. These latest results suggest that Rif1p acts either on the speed of the replication fork close to telomeres or on sub-telomeric origin firing, or both.

Personal contribution: 2D-gel analysis

By the mean of 2D-gel analysis, we plan to address the speed of the replication fork and the firing of new origins at telomeric sequences in presence or absence of Rif1. As preliminary data we reproduced data published in 2004 (Markovets et al. (2004) *MCB* **24**:4019-4031), looking at multiple copy origins located in the Y-telomeres (Fig43). Our results indicate that Rif1-depleted cells replicate the telomeric region 10-15min earlier than the control, but on the other hand undergo a stronger pausing at the repeats. This contradiction could be explained by the firing of close dormant origins. The next experiment will then be to analyse by 2D-gel the behaviour or these origins.

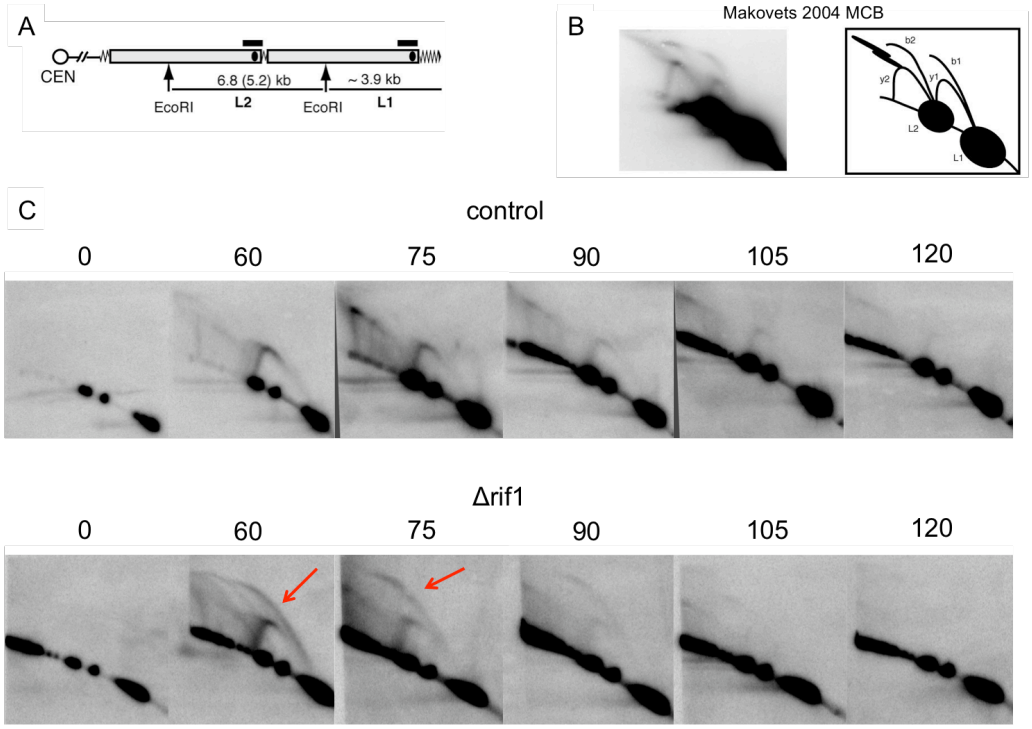


Figure 43. Replication intermediates of Y-telomeres detected by 2D gel electrophoresis. A. Schematic diagram of an EcoRI digest of a chromosome arm containing two tandem Y-elements (gray bars). Zigzag line represents TG1-3 tracts, black ellipsoids within analyzed DNA fragments show the position of the known ARSs, and black horizontal bars correspond to the fragments that were used as probes for the Southern hybridizations shown. B. Replication intermediates of Y-telomeres analyzed by 2D-gel (Makovets 2004). C. Reproduction of Makovets data by 2D-gel. Timecourse (time in min after alfa-F release): formation of bubble arc in Δrif1 samples shows that replication fork pauses at the telomeres.

5. Discussion I

The assays used so far to study replication and stability of specific genomic sequences *in vivo* in mammalian cells were limited to PCR or Southern blotting. Even studies on plasmids have been consistently limited by the amount of DNA recovered (in the range of nanogram), requiring bacterial amplification for molecular analysis of replication intermediates and products (Cleary et al., 2002).

This thesis describes a new method to obtain micrograms of plasmid DNA replicated *in vivo* in a single dish of cultured human cells, upon transfection and recovery of SV40 based constructs. The main modification to standard protocols lies in the expression of *large T antigen* (Tag) - the only viral factor required for SV40-based replication - directly from the replicating plasmid. In former protocols, these assays were performed in specific cell lines expressing the Tag from one/two copies of the gene integrated in the genome, thus producing a limited amount of protein. In our system we use plasmids containing the whole SV40 genome, allowing a positive feedback loop between protein expression and replication. Furthermore, the presence of Tag directly on the plasmids gives us the possibility to test their replication in any transfectable cell line. We also optimized a new protocol, adapted from the QIAgen miniprep procedure, in order to recover the plasmids and their RIs quickly (half a day), with high yield and low contamination of genomic DNA.

It is important to mention that we tested the safety of our plasmids, in that they are not competent to form viruses (data not shown). The reasons behind this are the following: 1) our plasmids are larger than 8kb, which is too big to be packed in the SV40 capsid; 2) the KpnI site, where we inserted the GAA repeats, is in a regulatory zone of SV40 coding for the encapsidation signal and the promoters of the capsid proteins, which impairs the packaging process.

The production and isolation of large amounts of plasmid DNA from human cells opened new possibilities for the *in vivo* analysis of replication intermediates. Taking advantage of these experimental conditions, we could visualize 2D-gel intermediates directly by ethidium bromide staining and extract specific areas of

the gel, in order to analyse by EM the corresponding DNA molecules. The combination of a) the high plasmid copy number, b) the low contamination of co-extracted genomic DNA and 3) the separation of different intermediates by 2D-gel electrophoresis, enabled us to obtain remarkably pure preparations of specific replication intermediates, greatly assisting their EM analysis. Besides confirming the established interpretation of standard 2D-gel signals, our experiments gave us the opportunity to uncover the molecular architecture of unusual intermediates, specifically associated with expanded GAA repeats.

Using the different methods described above, we characterized DNA replication across different numbers of GAA repeats. 2D-gel analyses showed that GAA repeats pause the fork in a length dependent manner (Fig15). The formation of GAA-specific intermediates start to be visible between 33 and 66 tracts, which correlates with the appearance of the RB in bacteria and the symptoms of the patients. Differently from common interpretations of published results, suggesting orientation dependent fork pausing at GAA repeats (bacteria see Pollard et al., 2004; yeast see Krasilnikova and Mirkin, 2004; Shishkin et al., 2009), we collected independent lines of evidence that expanded GAA repeats can pause replication forks in both orientations (Figs 15 to 17), although such pausing is differently manifested in the 2D gels. While several intermediates migrating on the Y-arc or in its close proximity identify forks paused at TTC repeats on the lagging strand, a single spot under the Y-arc marks fork pausing in the GAA orientation. On the other hand, GAA repeats on the lagging strand are associated with a stronger spot above the monomer spot, likely identifying processing events at GAA repeats behind the replication fork (see below). However, in both orientations, fork pausing appears to be transient, as supported independently by both EM mapping (Fig18) and 2D-gel data (Fig19).

A relevant and novel effect that GAA repeats exert on plasmid replication is the accumulation of symmetric X-molecules. In both orientations X intermediates accumulate with increasing numbers of repeats. We show that they represent two fully replicated plasmids joined at the repeats by a molecular junction that can involve only one strand of each molecule. The size of the connection in

certain molecules may suggest possible extensive annealing between these two strands. Interestingly, the formation of these GAA-specific X-structures strictly correlates with the disappearance of the 2n-spike signal. This 2n-spike signal is consistently observed during replication of control plasmids, as well as genomic DNA of different species (see Introduction). To understand if there is a possible link between these two phenomena we further characterized the molecules migrating as a spike. EM analyses revealed 2 fully replicated plasmids connected via a sequence homology-independent junction that links one single strand DNA of the first sister chromatid with the double helix of the second. Taken together these data strongly support a model in which replicated duplexes are often linked behind the forks by hemicatenation, as previously suggested (Lopes et al., 2003; Lucas and Hyrien, 2000), and not by recombination, fork reversal or termination. In some exceptional cases we could observe a replication fork and a postreplicative junction within the same molecule, confirming the hypothesis that hemicatenanes are formed post-replicatively and chase the forks in close proximity (Lopes et al., 2003).

The combination of these data allows us to propose a model - based on the one previously suggested by Lopes et al 2003 - to explain the conversion of transient sister-chromatid junctions into stable post-replicative intermediates at expanded GAA repeats (Fig44). During replication, hemicatenanes form behind the fork and freely branch migrate independently of sequence homology, leading to the formation of asymmetric X-molecules after plasmid replication and linearisation (Fig44 left). The presence of long GAA tracts may limit the free migration of the hemicatenanes and "align" the sister chromatids by annealing the repetitive sequences on the newly replicated strands and/or by forming triplex structures. We propose that this annealing would stabilize the junctions and possibly mediate, postreplicatively, the expansion mechanism (Fig44 right; see below).

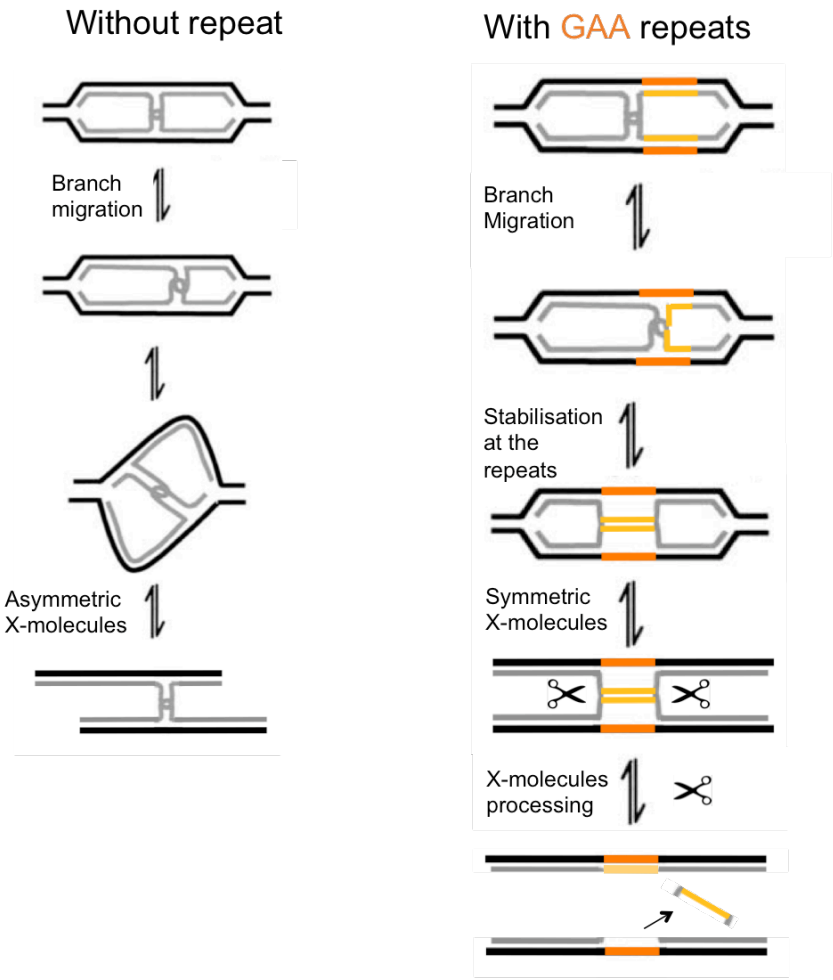


Figure 44. Model for the stabilisation of post-replicative junction at the repeats. In absence of repeats, post-replicative junctions form behind the fork and can freely branch migrate. In presence of repeats the post-replicative junction is stabilized by annealing of the two newly synthesized strands. The junction is then processed leading to the possible formation of linear DNA with single strand gap at the repeats (Modified from Lopes et al., 2003)

Gel extraction of the GAA/TTC specific intermediates revealed also the presence of fork reversal at the level of the repetitive tract. When GAA repeats are on the lagging strand template, the regressed arm is consistently ~300bp long, which correlates with the size of the repetitive sequence. This indicates that the fork reverses directly at the repeats or right after their replication. On the other hand, when TTC are on the lagging strand, the forks extracted from the pausing signals are indeed paused and reversed at the repeats, but display a more heterogenous size of the regressed arm, reaching up to ~1.75kb. Fork reversal was previously shown by EM to spontaneously form *in vitro* at repetitive DNA (Fouche et al., 2006), but its occurrence *in vivo* and its physiological relevance remained elusive. In our system, the difference in size of the regressed arms having GAA or

TTC repeats as lagging strand template suggests that this phenomenon happens *in vivo* by specific, orientation-dependent mechanisms.

Taking into consideration that stable post-replicative junctions and fork reversal are detected simultaneously on the GAA-carrying plasmids, we consider likely that both unusual intermediates originate from the same mechanism. The following model (Fig45) - which is slightly reminiscent of the one proposed by Liberi et al. 2005 to explain stable post-replicative junctions upon DNA damage - suggests that the hemicatenanes migrating behind the fork could induce the pairing of the two newly synthesized strands at the level of the repeats. In the GAA90 orientation (Fig45 left) the fork encountering the repeats would soon be converted into a reversed fork by the impact of the migrating hemicatenane with the fork paused at the repeats. Conversely, in TTC90 (Fig45 right) the fork would not reverse right after the repeats, but would keep progressing. Shortly after this, however, we suppose that the annealing formed at the repeats could induce further "unzipping" of synthesized strands, leading to long ssDNA behind the fork and finally to reversed forks with a markedly long regressed arm. Preliminary EM data (data not shown) seem to support this model, as we observed in TTC90 samples replication intermediates with very long single stranded stretches behind the fork and some unexplained double stranded arms. These intriguing results need to be followed up by further investigation.

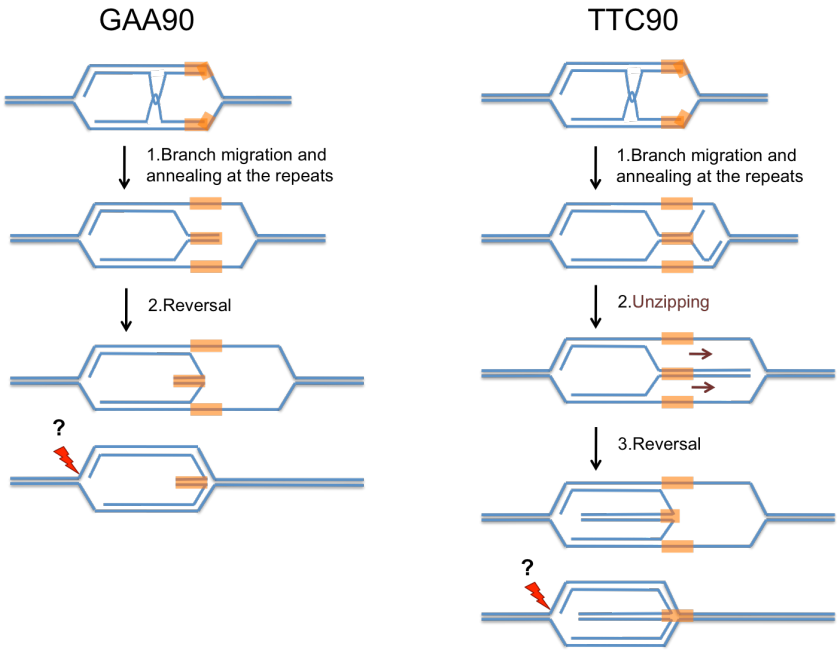


Figure 45. Model of hemicatenane-induced fork reversal at GAA90 or TTC90.

The majority of the regressed arms observed by EM were formed in RIs that we interpreted as broken bubbles. This interpretation is consistent with the data published in 1996 (Kalejta and Hamlin, 1996), which demonstrate that broken bubbles migrate as additional signals under the Y-arc (scheme fig46 black). However, the presence of a fourth regressed arm certainly modifies the migration of these molecules retarding them in both the 1st and the 2nd dimension. Relatively small "broken-reversed" bubbles (bubble size: 750-1500bp; regressed arm: up to 750bp) run in the spots detected under the Y-arc, as detected for GAA66/90 repeats (Fig46A). Conversely, the molecules with larger regressed and/or broken arms detected with TTC90 repeats (bubble size: 1750-3250bp; regressed arm: 1250-2250bp) tend to run between the Y-arc and the X spot (Fig46B).

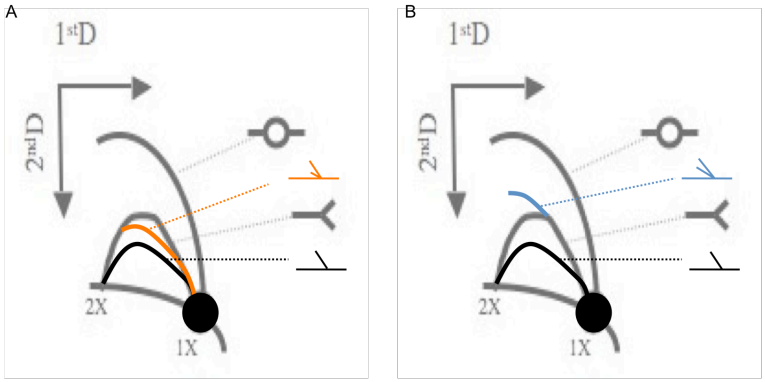


Figure 46. 2D-gel representation of broken bubbles (black) and broken bubbles with a regressed arm (orange and blue). A. Small broken bubbles with short regressed arm. B. Large broken bubble with large regressed arm.

It was rather surprising to realise, based on length measurements, that the fork undergoing breakage was not the one crossing the repeats, but rather the other one. This data suggests that a stalled replication fork can affect its sister, at least in this relatively small replicon on a plasmid. Delayed migration of the sister fork was also reported on genomic DNA, when the second fork in the bubble is encountering a double strand break on the template (Doksani et al., 2009). The mechanism by which this can happen remains unclear and will require further investigation. In this system, we hypothesise that extensive fork reversal detected at expanded GAA/TTC repeats may lead to a particularly inflexible fork structure, which could interfere with fork rotation and impair the redistribution of topological stress on the replicating plasmids by precatenane formation

(Postow et al., 2001). This accumulated stress may finally affect the integrity of the replication forks, particularly at the normal, "non-reversed fork" because of the presence of ssDNA, which is instead "protected" by extensive annealing at the reversed fork. Whether the fork breaks because of mechanical stress (*in vitro* or *in vivo*) or because of enzymatic processing will be one of the important biological questions to address in the future.

The last intermediates resulting from expanded GAA repeats are 1n molecules. They appear concomitantly with normal replication intermediates (data not shown) suggesting that they are also replication-dependent. Moreover, they contain single strand gaps consistently localized at the site of the repeats, as well as some flaps, which are instead more widespread. Formation of such gapped linear plasmids could be explained as result of processing of the stable post-replicative junctions formed at the repeats (see model Fig47).

We propose that resolution and formation of stable post-replicative junctions can be the cause of TNR instability. Fig47 describes two models, by which this mechanism could lead to GAA repeat expansion.

1) GAA expansion is induced by post-replicative junction processing (Fig47A).

   The resolution of the GAA specific post-replicative junctions by endonucleolytic cleavage can lead to the formation of a linear plasmid with a single strand gap at the repeats, which will then be refilled by DNA polymerase. A mispairing of the GAA repeats close to the available 3'-OH or a polymerase slippage event during post-replicative synthesis could lead to the formation of a loop, which will cause the expansion during gap filling.

2) Nick-dependent GAA expansion due to hemicatenane branch migration (Fig47B).

   As TNR pose a threat to polymerase activity, it is conceivable that DNA synthesis may be restarted by re-priming within the repeats, leaving transient discontinuities (nicks) on the newly replicated strand. Unwinding of nicked repeats by hemicatenane branch migration can lead

to GAA single strand flaps, which can form loops by inaccurate reannealing. This expansion mechanism can be coupled (Fig47B right) or not (Fig47B left) with the formation of stable post-replicative junction.

In the latter case (Fig47B left), branch migration in a nicked repetitive sequence can resolve the hemicatenane and induce the formation of a flap, which could be inaccurately reannealed to the template strand, forming a loop and finally leading to expansion.

In the former case (Fig47B right) looping of the 3'-OH end at a discontinuity could happen within a stable postreplicative junction during "template switch" synthesis (see Introduction) leaving a transient gap in front of its parental template. Resolution of the template switch intermediate and reannaealing of the elongated strand with its parental counterpart could lead to expansion.

These two models are in agreement with recent yeast data (Shishkin et al., 2009) showing that TNR instability does not in fact correlate with replication fork pausing, but is genetically linked to post-replicative repair pathways . We plan in the near future to perform a similar genetic analysis, in order to test our model and to assess whether these recent observations in yeast are reproducible in the mammalian system.

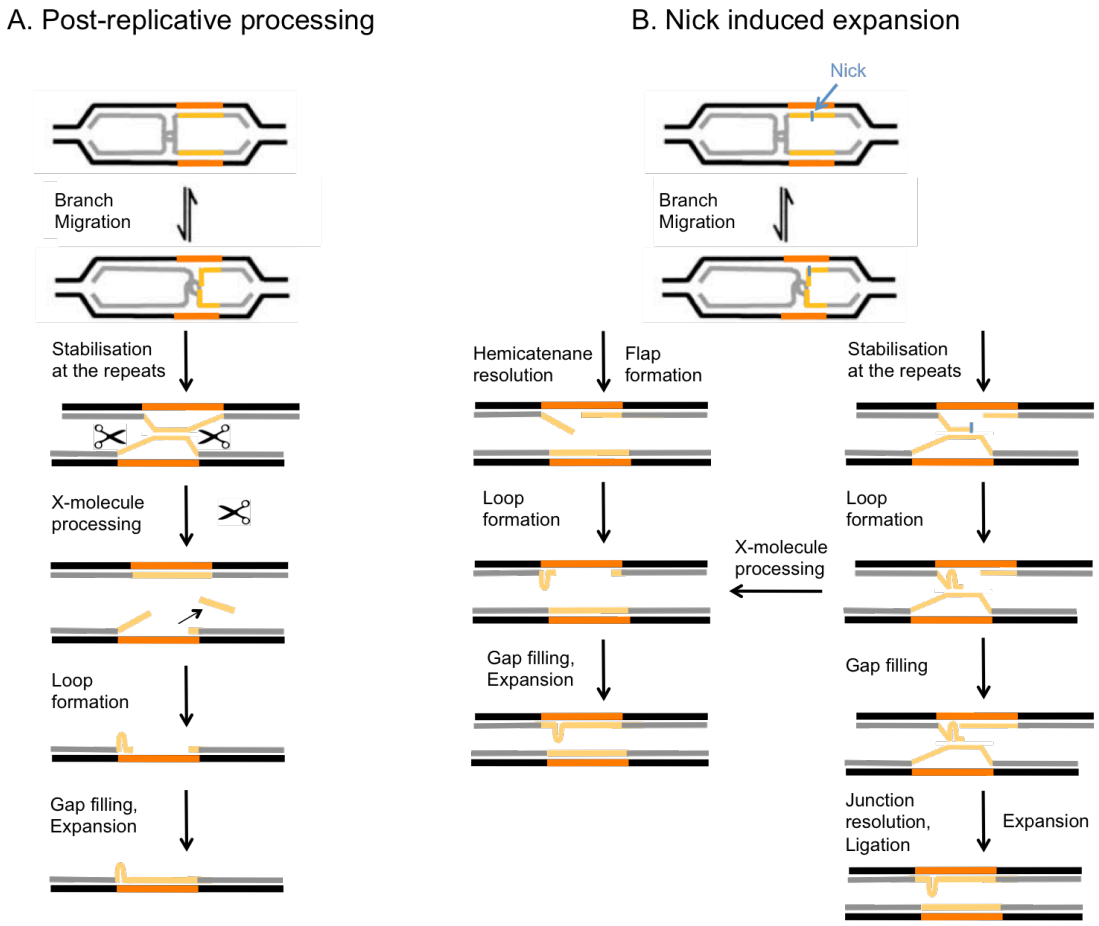


Figure 47. Expansion models. A. Repeat expansion due to the processing of post-replicative junction. B. Nick-dependent GAA expansion due to hemicatenane branch migration.

In conclusion, this thesis highlighed the interconnection between DNA replication and expanded GAA repeats *in vivo* in mammalian cells. We demonstrated the presence of post-replicative junctions behind the replication fork, which can be stabilised by long repetitive tracts and possibly contribute to their expansion. We also showed that GAA repeats can lead *in vivo* to fork reversal and fork breakage. These data allowed us to refine current models for TNR expansion, mostly attributed to post-replicative events.

6. Outlook

In the short term, we plan to further characterise the different X intermediates described in this thesis by denaturing EM protocols, in order to directly assess the presence of hemicatenanes. We also want to test if the GAA-specific intermediates described in this study can also form on a linear template or if it is specific for plasmid replication. To this purpose we will use yeast genomic DNA - where GAA tracts have been cloned close to an active replication origin - and analyse by 2D-gel the presence of specific post-replicative junctions and other GAA-specific RIs.

In the long run, our goal is to identify novel factors involved in GAA repeat instability. We plan to combine siRNA depletion of specific candidate genes with the 2D-gel analysis described here. This combined approach has already been optimized (data not shown) and will allow us to identify factors affecting the formation or resolution of GAA-specific intermediates. Once a candidate will be identified, its effects will be further caracterised by EM and by additional biochemical assays. Our first set of candidates will cover a set of helicases known to be involved in the resolution of secondary structures - like RecQ-helicases, Pif1 and RTEL1 - but also proteins involved in the replication pausing complex (Errico and Costanzo) and the post-replicative repair. We also plan to test different endonucleases and Holliday junction resolvases - like Mus81, Gen1, SLX1/4, XPF1 - to understand how the GAA-specific post-replicative junctions are finally resolved.
However, screening a high number of candidates is not feasible with a time-consuming approach as 2D gels. In order to shift to a high throughput analysis, we will consider different experimental set up that could rapidly detect TNR structure accumulation and instability, possibly taking advantage of our recently developed anti-RB antibody.

7. Discussion II

The data from Sakamoto 1999 (Sakamoto et al., 1999) showed for the first time that GAA repeats can induce inter- and intra-molecular junctions in *E. coli* in a length and orientation dependent manner. Further biochemical characterization of these junctions led the authors to propose RR*Y triplex as underlying structure in bacteria.

Our EM data demonstrate that GAA repeats can form a seemingly identical type of structure during *in vivo* replication in human cells. We also showed that these junctions are formed post-replicatively and that the junction can be limited to a single strand from each sister molecule, resembling standard DNA annealing. However, our technique does not have the potential to determine whether the junction is in fact a triplex, as the underlying Hoogsteen base pairing would probably be resolved during preparation of our EM samples. On the other hand, it will be possible to probe by EM the structural details of these junctions, by using different conditions for DNA spreading that allow a different degree of "denaturation" of the DNA molecules. In the most extreme conditions, we could potentially resolve the junctions into hemicatenanes, and assess whether this kind of structure is a possible precursor of the stable, homology-mediated junctions observed at GAA/TTC repeats .

The second part of this thesis focused on the establishment and characterization of a monoclonal antibody specific to GAA repeat-associated structures. As similar intermediates were observed in bacteria and human cells, we took advantage of the prokaryotic model to produce the antigen. With the collaboration of Toshio Mori's group we developed 3 antibodies specific for GAA secondary structure. Further characterization *in vitro* showed that only one antibody had a strong affinity and specificity for the antigen. However, even this antibody proved unable to recognize the structure when embedded into a large excess of competing DNA. These results unfortunately affect the possibility to use this precious tool for *in vivo* experiments, such as direct immunofluorescence (IF)-based detection of GAA-specific structures in the nucleus. Similarly these

observations seriously impair the design of a screen based on such IF read-out, as well as the development of a new diagnostic tool.

However, we do believe that this antibody can still have potential for revealing *in vitro* experiments. The *in vitro* transcription assay, which needs to be repeated with purified antibody (recently made available to the lab), could represent a proof-of-principle test to use GAA-specific antibody as a possible therapeutic approach to restore frataxin transcription. Combined with 2D-gel extraction, we could test by dot blotting which of the different intermediates we specifically detect contain the epitope recognized by the antibody. Finally, the antibody could also be used after blotting of agarose gels (2D-gels), once we find proper conditions to preserve the secondary structure along the procedure.

8. Material and Methods I

8.1. Plasmids

SV40 genomic DNA and pGEM plasmids were digested by EcoRI. pGEM was dephosphorylated with Shrimp alkaline phosphatase and ligated to SV40 genomic DNA using T4 DNA ligase overnight at 16°C (pML113). A MCS (KpnI-CCATATGGCGCGCTCCGGACATATGGGTAC-KpnI) was then introduced at the KpnI site of pML113 creating pML114 and pML115 according to the two possible orientations. The MCS contains two supplemental restriction sites, BspEI and BssHII.

pMP142 (GAA33), pMP178 (GAA66), pMP179 (GAA115) and pMP180 (GAA90) (published in (Ohshima et al., 1998) were digested by BspEI-BssHII in order to isolate the different repetitive tracts, which were then cloned into pML114 and pML115.

Spacer of the same size as the GAA90 insert was also inserted in pML113 plasmid at KpnI site. It was produced by PCR on pBLUEscript plasmid using GGGGTACCGAAAGCGAAAGGAGCGGG and CAAAAGCTGGGTACCGGG primers.

Table 2. Different plasmids used in this study

Number of repeats	GAA as lagging strand template	TTC as lagging strand template
0	pML113	
33	pML130	pML139
66	pML131	pML140
90	pML132	pML141
115	pML118	pML142
Spacer	pML135	

8.2. Plasmid replication in mammalian cells and extraction

293T or U2OS cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and Pen/Strep were transfected with the different plasmids using JetPRIME reagent from PolyPlus.

Plasmid DNA was recovered after 40h by a modified QIAgen miniprep protocol (Ziegler et al., 2004). Briefly, cells were resuspended in buffer P1, lysed with 0.66% SDS and incubated for 1h30 at 37°C with 0.5mg/ml Proteinase K. Then DNA was shortly denatured by 25mM NaOH for 1min, neutralized by buffer P3 and span down for 15min at 18000rpm. The supernatant was finally processed on miniprep columns as described in QIAprep Spin Miniprep protocol.

8.3. Bi-dimensional electrophoresis and hybridisation

Plasmids extracted from mammalian cells were digested by EcoRI-DpnI, EtOH precipitated and resuspended in TE buffer. 2D-gel electrophoresis was done as described in (Brewer and Fangman, 1987). First dimension was run on 0.4% agarose gel for 16h at 50V and second dimension was run on 1% agarose gel with EtBr for 8h30 at 140V (see Fig48). Gels were blotted by capillarity on BIORAD Zeta-probe membranes, which were then probed with SV40 genomic DNA or pGEM DNA. The resulting 2D-gel arcs are described in Fig49.

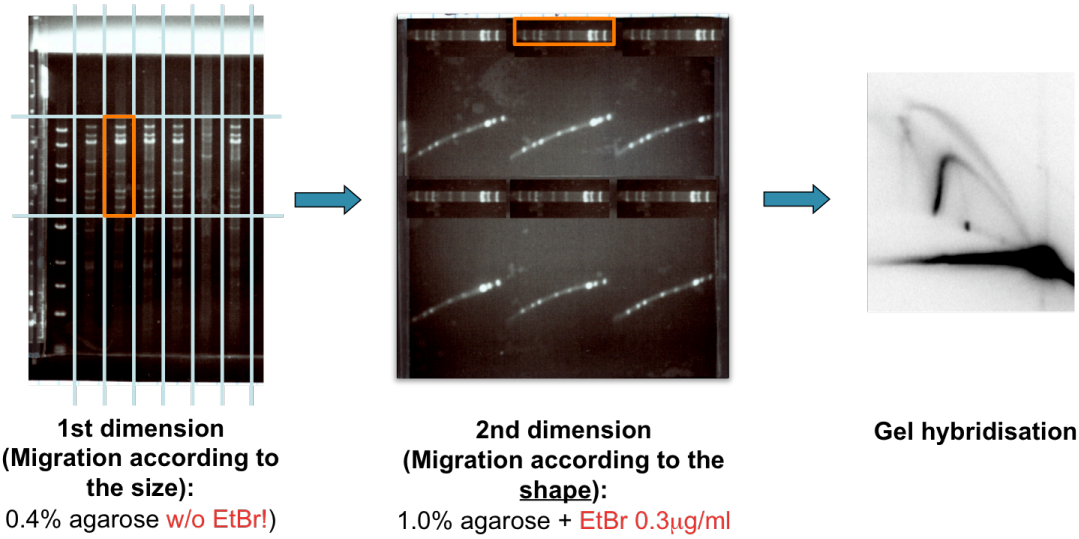


Figure 48. Description of neutral-neutral 2D-gel technique. Digested plasmids samples are first loaded on 0.4% agarose gel in absence of EtBr in order to favor a separation mostly according to size (1st D). Then the gel slice containing the RIs of interest is cut out and embedded in a new agarose gel for a second electrophoresis (2nd D). The second dimension contains EtBr in order to separate the RIs mostly according to their shape (bubble, Ys,...). Finally the gels are transferred by Southern blot and hybridized to reveal the RIs.

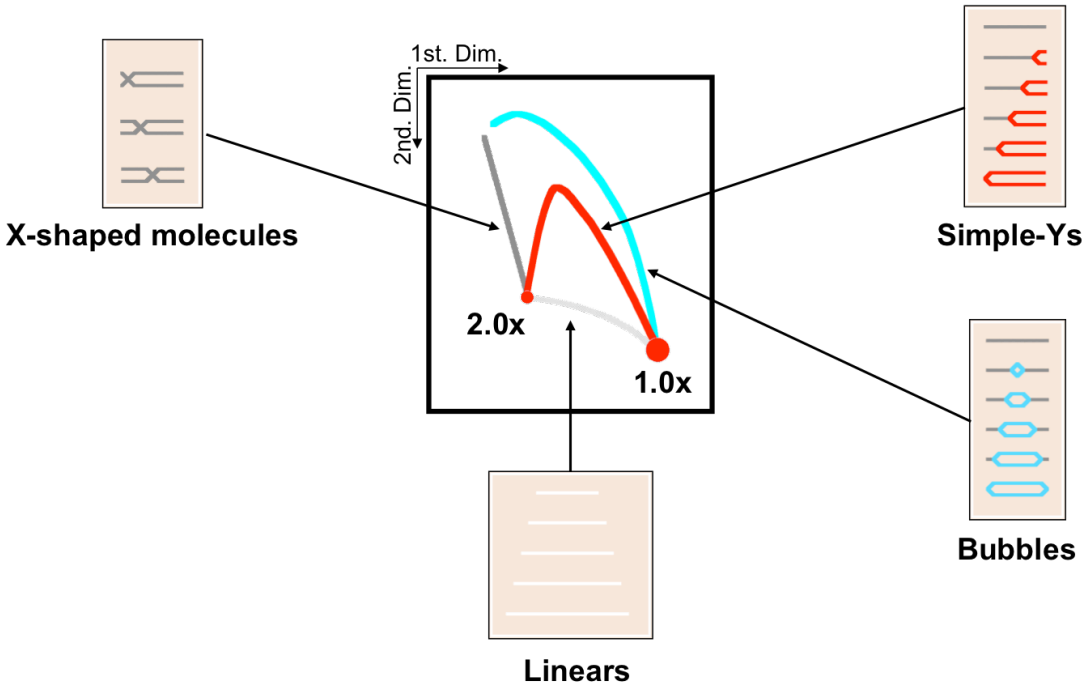


Figure 49. Description of the arcs resulting from a 2D-gel. Origin containing fragments display a typical pattern of divergent fork progression (bubble arc, blue). Fragments replicated passively give rise to single fork patterns (Y-arc, red). Recombination intermediates and other junctions between fully replicated fragments migrate in the 2n-spike (grey).

8.4. Electron microscopy and mapping of intermediates

Extracted DNA was digested with EcoRV, enriched for RIs on BND cellulose and processed for electron microscopy according to the protocol described in (Lopes, 2009).

RIs were measured using Image J. A color code was attributed to replicated (blue) and unreplicated (red) regions. The schematic representations of intermediates were then aligned from the least to the most replicated, arbitrarily positioning the origin of replication on the left side.

8.5. 2D-gel extraction for EM analysis

20-25ug of extracted DNA was processed by 2D-gel as described previously. After the 2nd dimension DNA intermediates were extracted from the gel and electro-eluted using the Elutrap electro-elution system from Whatman according to the instructions of the manufacturer. Recovered DNA was concentrated on Amicon Ultra centrifugal filters (Millipore) and analysed by EM.

9. Material and Methods II

9.1. GAA-based secondary structure enrichment

pMP179 plasmid containing GAA115 (Ohshima et al., 1998) was amplified in E.coli, linearised with XmnI and run on 0.8% agarose gel. The band retarded in the agarose gel (RB) was extracted and electro-eluted (Elutrap from Whatman). Finally the buffer was changed from TAE to PBS and concentrated on Amicon Ultra centrifugal filters (Millipore).

9.2. Monoclonal antibody production (Toshio Mori's laboratory)

RB was conjugated with BSA and the mice were immunized as described in Iwamoto et al., 2004. The mouse spleens were recovered, dissociated and fused with myeloma cells to form hybridomas. Finally, hybridomas were screened by ELISA for the identification of specific antibodies. Supernatants from clones C23, N49 and T57 were identified as RB specific antibody and then concentrated to a factor 1:20, via lyophilisation.

9.3. Immuno-fluorescence

U2OS cells were plated on coverslips and cultured for 24hrs in the growth medium (DMEM + 10% FBS). Plasmids were transfected using jetPRIME reagent. After 16h or 48h, cells were washed once with 1xPBS, fixed with pre-cooled MetOH for 30min at -20°C, washed with 1xPBS and fixed again with 4% formalin for 10 min at room temperature. Cells were then permeabilized with 0.5% Triton-X 100 in 1xPBS for 5 min on ice. After washing with 1xPBS, cells were blocked in 20% FBS (fetal bovine serum) for 30min at 37 °C with mild shaking. Coverslips were then washed with 1xPBS and incubated o/n at 4°C with the monoclonal antibody against GAA-RB [C23, N49, T57]. The next day cells were washed with 1xPBS and incubated with goat anti-mouse Alexa Fluor 488 IgG (1/50; Molecular Probes) for 30 min at 37 °C with mild shaking.

After washing with 1xPBS, nuclear DNA was counterstained with DAPI (0.5 μg/ml PBS) for 5 min at room temperature. Cells were then washed with 1xPBS, the dishes were mounted in drops of Vectashield and sealed with coverslips. Then, fluorescence microscopy was performed using a Leica DMIRB microscope.

Fluorescent images of RB and of DNA were processed using Adobe Photoshop software.

Different fixation conditions also tested:

1) 4% formalin only (10min at RT)
2) MetOH only (30min at -20°C)
3) With pre-extraction:
   - pre-extraction: 25mM Hepes 7.4, 50mM NaCl, 1mM EDTA, 3mM $MgCl_2$, 300mM Sucrose, 0.5% TritonX-100 for 10min on ice.
   - 1x wash with 1xPBS
   - Fixation either with 4% Formaldehyde (at RT for 10') and/or MetOH (30' at -20°C)

9.4. Dot blot and Immuno-detection

DNA was loaded onto dot-blot apparatus and blotted on BIORAD Zeta-probe membrane. The membrane was washed 2x with 2xSSC for 10min and 1x with 1xPBS. Then it was processed for immuno-detection using the following conditions: 1h of blocking in 2% ECL Prime blocking agent (GE Healthcare) in 1xPBS, overnight incubation of lyophilised DNA antibody (1:400) or lyophilised RB antibody (1:500) at 4°C, 2 washes with 1xPBS for 20min, 1 wash in 1xPBS 0.1% Tween for 20min, 1h incubation of anti-mouse antibody (1:5000) at RT, 3 washes in 1xPBS 0.1% Tween for 15min. Signal was revealed using Amersham ECL™ Western Blotting Detection Reagents

9.5. Formation of secondary structures *in vitro* using oligonucleotides

Secondary structures are formed by the annealing of the oligonucleotides (oligos) as mentioned in Table 3. The specific oligos (100nM) were mixed in endonuclease buffer (1:10) (250mM HEPES-KOH pH 7.6, 1M KCl, 20mM MgCl2), incubated at 95°C for 10min and allowed to cool down over night.

Table 3. Sequence of the oligos used for secondary structure formation.

Oligos	Sequence
BubTop	5'-CCAGTGATCACATACGCTTTGCTATTCCGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT CCGTGCCACGTTGTATGCCCACGTTGCCG-3'
BubBot	5'-CGGTCAACGTGGGCATACAACGTGGCACGGGAGCGTTAACCCTAACCCTAACCCTAACCC CCGGAATAGCAAAGCGTATGTGATCACTGG-3'
5'-Inv	5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGGTTAGGGTTAGGGTTAGGGTTAACGCTC-3'
3'-Inv	5'-GGGTTAGGGTTAGGGTTAGGGTTAACGCTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-3'
F7 (30 mer)	5'-ATTGACTAGGTTACATGACTGAATGATAGT-3
F8 (30 mer)	5'-GGAGTAAAGTACTAGGTATGTCGACATTGA-3'
F9 (59 mer)	5'-ACTATCATTCAGTCATGTAACCTAGTCAATCTGCGAGCTCGAATTCACTGGAGTGACCT-3'
F10 (60 mer)	5'-GAGGTCACTCCAGTGAATTCGAGCTCGCAGTCAATGTCGACATACCTAGTACTTTACTCC-3'
Stem	5'-GCCAGCGCTCGGTTTTTTTTTTTTTTTTTTTTTTCCGAGCGCTGGC-3'
X12-1	5'-GACGCTGCCGAATTCTGGCTTGCTAGGACATCTTTGCCCACGTTGACCCG- 3'
X12-2	5'-CGGGTCAACGTGGGCAAAGATGTCCTAGCAATGTAATCGTCTATGACGTC -3'
X12-3	5'- GACGTCATAGACGATTACATTGCTAGGACATGCTGTCTAGAGACTATCGC -3'
X12-4	5'- GCGATAGTCTCTAGACAGCATGTCCTAGCAAGCCAGAATTCGGCAGCGTC -3'

Table 4. Oligos required to form *in vitro* the different secondary structures.

Name	Annealed oligos
Splayed arm	F9, F10
3'-flap	F7, F9, F10
5'-flap	F8, F9, F10
3-way junction	F7, F8, F9, F10
Stem loop	Stem
5' D-loop	BubTop, BubBot and 5'-Inv.
3' D-loop	BubTop, BubBot and 3'-Inv
Holliday junction	X12-1, X12-2, X12-3, X12-4

9.6. Plasmids

Plasmids (GAA0, GAA11, GAA44, GAA88) used for *in vitro* transcription assay were provided by Dr. Ed Grabczyk (Grabczyk and Usdin, 2000b).

9.7. *In vitro* transcription assay

The *in vitro* transcription reactions were processed using HiScribe™ T7 In Vitro Transcription Kit from NEB. 200ng of template DNA was used in a 20ul final volume and incubated at 42°C for 20min. Reaction was stopped by adding 20ul of 3M sodium acetate pH5.2 in 160ul of ddH_2O. DNA/RNA was finally purified by phenol/chlorophorm, resuspended in RNA loading dye (96 % Formamide, 10mM EDTA, 0.05 % Bromophenol Blue, 0.05 % Xylene Cyanole) and loaded on 1% agarose gel.

For some reactions, different concentrations of RB antibody were added to the reaction.

10. References

Agazie, Y.M., Lee, J.S., and Burkholder, G.D. (1994). Characterization of a new monoclonal antibody to triplex DNA and immunofluorescent staining of mammalian chromosomes. J Biol Chem *269*, 7019-7023.

Alberts, B. (2003). DNA replication and recombination. Nature *421*, 431-435.

Atkinson, J., and McGlynn, P. (2009). Replication fork reversal and the maintenance of genome stability. Nucleic Acids Res *37*, 3475-3492.

Balakrishnan, L., Gloor, J.W., and Bambara, R.A. (2010). Reconstitution of eukaryotic lagging strand DNA replication. Methods *51*, 347-357.

Baralle, M., Pastor, T., Bussani, E., and Pagani, F. (2008). Influence of Friedreich ataxia GAA noncoding repeat expansions on pre-mRNA processing. Am J Hum Genet *83*, 77-88.

Benard, M., Maric, C., and Pierron, G. (2001). DNA replication-dependent formation of joint DNA molecules in Physarum polycephalum. Mol Cell *7*, 971-980.

Blow, J.J., and Dutta, A. (2005). Preventing re-replication of chromosomal DNA. Nature reviews Molecular cell biology *6*, 476-486.

Blow, J.J., and Hodgson, B. (2002). Replication licensing--defining the proliferative state? Trends Cell Biol *12*, 72-78.

Branzei, D. Ubiquitin family modifications and template switching. (2011). FEBS Lett *585*, 2810-2817.

Branzei, D., and Foiani, M. Maintaining genome stability at the replication fork. (2010). Nat Rev Mol Cell Biol *11*, 208-219.

Branzei, D., Vanoli, F., and Foiani, M. (2008). SUMOylation regulates Rad18-mediated template switch. Nature *456*, 915-920.

Brewer, B.J., and Fangman, W.L. (1987). The localization of replication origins on ARS plasmids in S. cerevisiae. Cell *51*, 463-471.

Broker, T.R., and Lehman, I.R. (1971). Branched DNA molecules: intermediates in T4 recombination. J Mol Biol *60*, 131-149.

Chavez, A., Tsou, A.M., and Johnson, F.B. (2009). Telomeres do the (un)twist: helicase actions at chromosome termini. Biochim Biophys Acta *1792*, 329-340.

Chen, X., Mariappan, S.V., Catasti, P., Ratliff, R., Moyzis, R.K., Laayoun, A., Smith, S.S., Bradbury, E.M., and Gupta, G. (1995). Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological implications. Proceedings of the National Academy of Sciences of the United States of America *92*, 5199-5203.

Chou, D.M., and Elledge, S.J. (2006). Tipin and Timeless form a mutually protective complex required for genotoxic stress resistance and checkpoint function. Proc Natl Acad Sci U S A *103*, 18143-18147.

Cleary, J.D., Nichol, K., Wang, Y.H., and Pearson, C.E. (2002). Evidence of cis-acting factors in replication-mediated trinucleotide repeat instability in primate cells. Nat Genet *31*, 37-46.

Daee, D.L., Mertz, T., and Lahue, R.S. (2007). Postreplication repair inhibits CAG.CTG repeat expansions in Saccharomyces cerevisiae. Mol Cell Biol *27*, 102-110.

Ditch, S., Sammarco, M.C., Banerjee, A., and Grabczyk, E. (2009). Progressive GAA.TTC repeat expansion in human cell lines. PLoS genetics *5*, e1000704.

Doksani, Y., Bermejo, R., Fiorani, S., Haber, J.E., and Foiani, M. (2009). Replicon dynamics, dormant origin firing, and terminal fork integrity after double-strand break formation. Cell *137*, 247-258.

Errico, A., and Costanzo, V. (2010). Differences in the DNA replication of unicellular eukaryotes and metazoans: known unknowns. EMBO Rep *11*, 270-278.

Flores, M.J., Bierne, H., Ehrlich, S.D., and Michel, B. (2001). Impairment of lagging strand synthesis triggers the formation of a RuvABC substrate at replication forks. EMBO J *20*, 619-629.

Fouche, N., Ozgur, S., Roy, D., and Griffith, J.D. (2006). Replication fork regression in repetitive DNAs. Nucleic Acids Res *34*, 6044-6050.

Fousteri, M., and Mullenders, L.H. (2008). Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. Cell research *18*, 73-84.

Frank-Kamenetskii, M.D., and Mirkin, S.M. (1995). Triplex DNA structures. Annu Rev Biochem *64*, 65-95.

Fry, M., and Loeb, L.A. (1994). The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. Proc Natl Acad Sci U S A *91*, 4950-4954.

Grabczyk, E., Mancuso, M., and Sammarco, M.C. (2007). A persistent RNA.DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP in vitro. Nucleic acids research *35*, 5351-5359.

Grabczyk, E., and Usdin, K. (2000a). Alleviating transcript insufficiency caused by Friedreich's ataxia triplet repeats. Nucleic acids research *28*, 4930-4937.

Grabczyk, E., and Usdin, K. (2000b). The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. Nucleic acids research *28*, 2815-2822.

Grant, L., Sun, J., Xu, H., Subramony, S.H., Chaires, J.B., and Hebert, M.D. (2006). Rational selection of small molecules that increase transcription through the GAA repeats found in Friedreich's ataxia. FEBS Lett *580*, 5399-5405.

Hanvey, J.C., Shimizu, M., and Wells, R.D. (1988). Intramolecular DNA triplexes in supercoiled plasmids. Proc Natl Acad Sci U S A *85*, 6292-6296.

Higgins, N.P., Kato, K., and Strauss, B. (1976). A model for replication repair in mammalian cells. Journal of molecular biology *101*, 417-425.

Huppert, J.L., and Balasubramanian, S. (2007). G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res *35*, 406-413.

Huppert, J.L., Bugaut, A., Kumari, S., and Balasubramanian, S. (2008). G-quadruplexes: the beginning and end of UTRs. Nucleic Acids Res *36*, 6260-6268.

Iwamoto, T.A., Kobayashi, N., Imoto, K., Yamamoto, A., Nakamura, Y., Yamauchi, Y., Okumura, H., Tanaka, A., Hanaoka, F., Shibutani, S.*, et al.* (2004). In situ detection of acetylaminofluorene-DNA adducts in human cells using monoclonal antibodies. DNA Repair (Amst) *3*, 1475-1482.

Kalejta, R.F., and Hamlin, J.L. (1996). Composite patterns in neutral/neutral two-dimensional gels demonstrate inefficient replication origin usage. Mol Cell Biol *16*, 4915-4922.

Kim, H.M., Narayanan, V., Mieczkowski, P.A., Petes, T.D., Krasilnikova, M.M., Mirkin, S.M., and Lobachev, K.S. (2008). Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. EMBO J *27*, 2896-2906.

Koster, D.A., Palle, K., Bot, E.S., Bjornsti, M.A., and Dekker, N.H. (2007). Antitumour drugs impede DNA uncoiling by topoisomerase I. Nature *448*, 213-217.

Kovtun, I.V., and McMurray, C.T. (2008). Features of trinucleotide repeat instability in vivo. Cell Res *18*, 198-213.

Krasilnikova, M.M., and Mirkin, S.M. (2004). Replication stalling at Friedreich's ataxia (GAA)n repeats in vivo. Molecular and cellular biology *24*, 2286-2295.

Law, M.J., Lower, K.M., Voon, H.P., Hughes, J.R., Garrick, D., Viprakasit, V., Mitson, M., De Gobbi, M., Marra, M., Morris, A.*, et al.* (2010). ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. Cell *143*, 367-378.

Lee, J.S., Burkholder, G.D., Latimer, L.J., Haug, B.L., and Braun, R.P. (1987). A monoclonal antibody to triplex DNA binds to eucaryotic chromosomes. Nucleic Acids Res *15*, 1047-1061.

Liberi, G., Maffioletti, G., Lucca, C., Chiolo, I., Baryshnikova, A., Cotta-Ramusino, C., Lopes, M., Pellicioli, A., Haber, J.E., and Foiani, M. (2005). Rad51-dependent DNA structures accumulate at damaged replication forks in sgs1 mutants defective in the yeast ortholog of BLM RecQ helicase. Genes Dev *19*, 339-350.

Lin, Y., and Wilson, J.H. (2007). Transcription-induced CAG repeat contraction in human cells is mediated in part by transcription-coupled nucleotide excision repair. Molecular and cellular biology *27*, 6209-6217.

Lopes, M. (2009). Electron microscopy methods for studying in vivo DNA replication intermediates. Methods Mol Biol *521*, 605-631.

Lopes, M., Cotta-Ramusino, C., Liberi, G., and Foiani, M. (2003). Branch migrating sister chromatid junctions form at replication origins through Rad51/Rad52-independent mechanisms. Mol Cell *12*, 1499-1510.

Lopez Castel, A., Cleary, J.D., and Pearson, C.E. (2010). Repeat instability as the basis for human diseases and as a potential target for therapy. Nature reviews Molecular cell biology *11*, 165-170.

Lucas, I., and Hyrien, O. (2000). Hemicatenanes form upon inhibition of DNA replication. Nucleic Acids Res *28*, 2187-2193.

Masamune, Y., and Richardson, C.C. (1971). Strand displacement during deoxyribonucleic acid synthesis at single strand breaks. J Biol Chem *246*, 2692-2701.

McLennan, Y., Polussa, J., Tassone, F., and Hagerman, R. Fragile x syndrome. (2011). Curr Genomics *12*, 216-224.

McMurray, C.T. Mechanisms of trinucleotide repeat instability during human development. (2010). Nat Rev Genet *11*, 786-799.

Mirkin, E.V., and Mirkin, S.M. (2007). Replication fork stalling at natural impediments. Microbiol Mol Biol Rev *71*, 13-35.

Ohno, M., Fukagawa, T., Lee, J.S., and Ikemura, T. (2002). Triplex-forming DNAs in the human interphase nucleus visualized in situ by polypurine/polypyrimidine DNA probes and antitriplex antibodies. Chromosoma *111*, 201-213.

Ohshima, K., Montermini, L., Wells, R.D., and Pandolfo, M. (1998). Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication in vivo. J Biol Chem *273*, 14588-14595.

Pandolfo, M. (2009). Friedreich ataxia: the clinical picture. J Neurol *256 Suppl 1*, 3-8.

Pearson, C.E., Wang, Y.H., Griffith, J.D., and Sinden, R.R. (1998). Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG)n. (CAG)n repeats from the myotonic dystrophy locus. Nucleic Acids Res *26*, 816-823.

Pollard, L.M., Sharma, R., Gomez, M., Shah, S., Delatycki, M.B., Pianese, L., Monticelli, A., Keats, B.J., and Bidichandani, S.I. (2004). Replication-mediated instability of the GAA triplet repeat mutation in Friedreich ataxia. Nucleic Acids Res *32*, 5962-5971.

Postow, L., Crisona, N.J., Peter, B.J., Hardy, C.D., and Cozzarelli, N.R. (2001). Topological challenges to DNA replication: conformations at the fork. Proc Natl Acad Sci U S A *98*, 8219-8226.

Punga, T., and Buhler, M. (2010). Long intronic GAA repeats causing Friedreich ataxia impede transcription elongation. EMBO Mol Med *2*, 120-129.

Pursell, Z.F., Isoz, I., Lundstrom, E.B., Johansson, E., and Kunkel, T.A. (2007). Yeast DNA polymerase epsilon participates in leading-strand DNA replication. Science *317*, 127-130.

Rindler, P.M., and Bidichandani, S.I. (2011). Role of transcript and interplay between transcription and replication in triplet-repeat instability in mammalian cells. Nucleic Acids Res *39*, 526-535.

Sakamoto, N., Chastain, P.D., Parniewski, P., Ohshima, K., Pandolfo, M., Griffith, J.D., and Wells, R.D. (1999). Sticky DNA: self-association properties of long

GAA.TTC repeats in R.R.Y triplex structures from Friedreich's ataxia. Mol Cell *3*, 465-475.

Sampathi, S., and Chai, W. (2011). Telomere replication: poised but puzzling. J Cell Mol Med *15*, 3-13.

Santos, R., Lefevre, S., Sliwa, D., Seguin, A., Camadro, J.M., and Lesuisse, E. (2010). Friedreich ataxia: molecular mechanisms, redox considerations, and therapeutic opportunities. Antioxid Redox Signal *13*, 651-690.

Segurado, M., Gomez, M., and Antequera, F. (2002). Increased recombination intermediates and homologous integration hot spots at DNA replication origins. Mol Cell *10*, 907-916.

Sercin, O., and Kemp, M.G. (2011). Characterization of functional domains in human Claspin. Cell Cycle *10*, 1599-1606.

Shishkin, A.A., Voineagu, I., Matera, R., Cherng, N., Chernet, B.T., Krasilnikova, M.M., Narayanan, V., Lobachev, K.S., and Mirkin, S.M. (2009). Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. Mol Cell *35*, 82-92.

Sogo, J.M., Lopes, M., and Foiani, M. (2002). Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. Science *297*, 599-602.

Soragni, E., Herman, D., Dent, S.Y., Gottesfeld, J.M., Wells, R.D., and Napierala, M. (2008). Long intronic GAA*TTC repeats induce epigenetic changes and reporter gene silencing in a molecular model of Friedreich ataxia. Nucleic Acids Res *36*, 6056-6065.

Tanaka, K. Multiple functions of the S-phase checkpoint mediator. (2010). Biosci Biotechnol Biochem *74*, 2367-2373.

Vetcher, A.A., Napierala, M., and Wells, R.D. (2002). Sticky DNA: effect of the polypurine.polypyrimidine sequence. J Biol Chem *277*, 39228-39234.

Vilchez, R.A., and Butel, J.S. (2004). Emergent human pathogen simian virus 40 and its role in cancer. Clin Microbiol Rev *17*, 495-508, table of contents.

Voineagu, I., Surka, C.F., Shishkin, A.A., Krasilnikova, M.M., and Mirkin, S.M. (2009). Replisome stalling and stabilization at CGG repeats, which are responsible for chromosomal fragility. Nat Struct Mol Biol *16*, 226-228.

Ziegler, K., Bui, T., Frisque, R.J., Grandinetti, A., and Nerurkar, V.R. (2004). A rapid in vitro polyomavirus DNA replication assay. J Virol Methods *122*, 123-127.

Zuccato, C., Valenza, M., and Cattaneo, E. Molecular mechanisms and potential therapeutical targets in Huntington's disease. (2010). Physiol Rev *90*, 905-981.

11. Acknowledgments

I would like to specially thank Massimo Lopes, for giving me the opportunity to do my PhD in his lab. He was an amazing supervisor, who always managed to transmit his enthusiasm for science and for fun.

Thank you to Prof. Toshio Mori, a great collaborator and also a member of my committee meeting, who plans to travel from Japan to attend my thesis defense.

Thank you also to all the European members of my committee meeting, Joachim Lingner, Peter Beard and Urs Greber, for their availability and helpful inputs.

I'm grateful to the members of my lab, Arnab, Judith, Isabella, Akshay, Raquel and Kai, for their help, but mainly for the nice time we had together. A particular thank you to Judith, who contributed to my project during her master and did a very good job and to Arnab, whose friendship and scientific knowledge were of a great support during the last 4 years. Thank you also to Kai for the German translation.

Last, I want to acknowledge my parents, my sister and her husband, for always being on my side in the good as well as in the difficult periods I had to face.

Curriculum Vitae

Surname: **Follonier**

Name: **Cindy**

Date of birth: **27.02.1984**

Place of birth: **Vernamiège VS**

Nationality: **Swiss**

Education:

2008/2012:

PhD at the Institute of Molecular Cancer Research. University of Zurich (Switzerland). Group of Prof. Massimo Lopes.
"Structural analysis of DNA replication across unstable repetitive sequences".

2006/2007:

Master in biology at the University of Geneva (Switzerland). Group of Master project in the group of Prof. David Shore.
"Identification of proteins capping the long telomeres using a screen".

2005/2006:

Last year of Bachelor in biology at the University of Manchester (UK). Bachelor project in the group of Prof. Andrew Sharrocks.
"Overexpression of key domains in the Forkhead transcription factor FoxK1".

2003/2005:

1st and 2nd year at the University of Geneva (Switzerland).
Bachelor monograph in the group of Prof. David Shore (2004/2005).
"L'influence du stress sur les télomères".

2002/2003:

"Maturité Fédérale" (final certificate of secondary education) at "Lycée Collège des Creusets" in Sion (VS/Switzerland).

Professional experience:

Summer 2006:

Summer undergraduate program at ISREC (Swiss institute of experimental research in cancers) in Lausanne (VD/Switzerland) in the group of Prof. Joachim Lingner.

Publications:

Giannattasio M, Follonier C, Tourrière H, Puddu F, Lazzaro F, Pasero P, Lopes M, Plevani P, Muzi-Falconi M. (2010) Exo1 competes with repair synthesis, converts NER intermediates to long ssDNA gaps, and promotes checkpoint activation. *Molecular Cell* **40**:1-13